

Welcome to the 16th Annual CALDER Research Conference

CALDERCENTER.ORG



Thank You!



WALTON FAMILY
FOUNDATION



Institute of Education
Sciences

Today's Agenda

Time	Session
9:00 a.m.—9:15 a.m.	Welcome & Announcements
9:15 a.m.—9:30 a.m.	Coming Attractions
9:30 a.m.—11:15 a.m.	Session I: COVID's Long Shadow, Five Years Later
11:15 a.m.—12:45 p.m.	Session II: Predicting Students' Post-secondary Success
12:45 p.m.—1:30 p.m.	Lunch
1:30 p.m.—3:00 p.m.	Session III: The Future of Annual Testing—Roundtable Discussion

CALDER Recruitment & Retention (CALDER R&R)

IES Funded Research and Development Center

Many researchers, many projects: 10 institutions across 9 states + metro Atlanta area

Team

AIR	Partners	
Dan Goldhaber (PI)	David Blazar, University of Maryland	Melissa Arnold Lyon, University at Albany SUNY
Ben Backes (Co-PI)	Sade Bonilla, University of Pennsylvania	Josh McGee, University of Arkansas
James Cowan (Co-PI)	Meagan Comb, Boston University	Ben Ost, University of Illinois Chicago
Michael DeArmond (Co-PI)	Kevin Fortner, Georgia State University	Steve Rivkin, University of Illinois Chicago
Emily Morton (Co-PI)	Thomas Goldring, Georgia Policy Labs	Matthew Springer, Basis Policy Research
Roddy Theobald (Co-PI)	Eric Hanushek, Stanford University	Elizabeth Steiner, RAND
Keane Alavi	Scott Imberman, Michigan State University	Gema Zamarro, University of Arkansas
Audrey Altieri	Tara Kilbride, Michigan State University	
Lisa Lachlan-Hache	Cory Koedel, University of Missouri	
Melia Repko-Erwin	Constance Lindsey, University of North Carolina	

Two Foundational Perspectives on the Teacher Labor Market

- **Understanding Teacher Shortages**

- Not a universal crisis, but targeted challenges:
 - » Subject areas
 - » Demographics
 - » School types
 - » Geographic regions

- **Importance of the Teacher Pipeline**

- The path from preparation to the classroom shapes the teacher workforce
 - » Teacher education programs and training
 - » Licensure
 - » Hiring processes
 - » Working conditions & compensation
 - » Retention efforts

Investigating Policies in Five Buckets

- **New Recruitment and Pathways**
 - Grow-your-own initiatives; grants for training; diversification initiatives
- **In-Preparation**
 - Stipends to support for teacher training
 - Labor market information for prospective teachers in college
- **Licensure**
 - Reduced licensure requirements
- **Compensation**
 - Statewide minimum salaries
 - Performance Pay
 - Targeted Bonuses
- **Working Conditions & Teacher Roles**
 - Four-day school week
 - Multi-classroom teachers
 - Work rules (CBAs)

~5% for TBD projects

Spotlight: In-Preparation Information Experiment

- Finding a teaching job can be easier (or harder) depending on the subject area/endorsement.
 - Supply exceeds demand for some positions (e.g., elementary teachers) but not others (e.g., special education, science, math...).
- Does giving undergraduate students information about the job prospects of newly licensed teachers *in different subject/endorsement areas* alter their views about teaching? Does it affect their subsequent choice of major?
- Experiments in 13 teacher education programs, including...



Session I: COVID's Long Shadow, Five Years Later

Paper Presentations

- **Supply and Demand in the Post-Pandemic Teacher Labor Market.** Roddy Theobald, CALDER/AIR
- **The Decline in Teacher Workforce Conditions During and After the COVID Pandemic.** Cory Koedel, University of Missouri, CALDER
- **What Do We Know About Academic Recovery Interventions?** Emily Morton, CALDER/AIR
- **ESSER and Student Achievement: Assessing the Impacts of the Largest One-time Federal Investment in K12 Schools.** Dan Goldhaber, CALDER/AIR

Discussant

- **Kevin Mahnken, The 74**

Supply and Demand in the Post-Pandemic Teacher Labor Market

Roddy Theobald, CALDER/AIR

Dan Goldhaber, CALDER/AIR & University of Washington

Grace Falken, University of Washington

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C240007 to American Institutes for Research. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

[CALDERCENTER.ORG](https://caldercenter.org)



Motivation

What is the state of the postpandemic teacher labor market? Is the supply of new teachers keeping up with demand?

Supply < Demand?

- Attrition is up so demand may be up (e.g., Bacher-Hicks et al., 2023; Camp et al., 2023; Goldhaber & Theobald, 2023)
- Worsening working conditions may further increase demand and decrease supply (e.g., Doan et al., 2023; Kraft et al., 2021)
- Supply may drop as emergency license policies end (e.g., Backes et al., 2024)

Supply > Demand?

- Student enrollment is down so demand for teachers may be down as well (e.g., Schueler & Miller, 2024)
- End of ESSER may lead to job cuts, thus decreasing demand (Goldhaber et al., 2024)
- Number of program completers has increased nationally, so supply may be up (Kraft & Lyon, 2024)

Perhaps most importantly, these trends likely vary across different teacher specialty areas like STEM and special education (e.g., Edwards et al., 2024; Mason-Williams et al., 2023)

Research Questions

1. How have supply and demand for teachers changed relative to prepandemic years?
2. How does postpandemic demand for teachers differ across specific subject areas and relative to the recent supply of new teachers in each subject area?

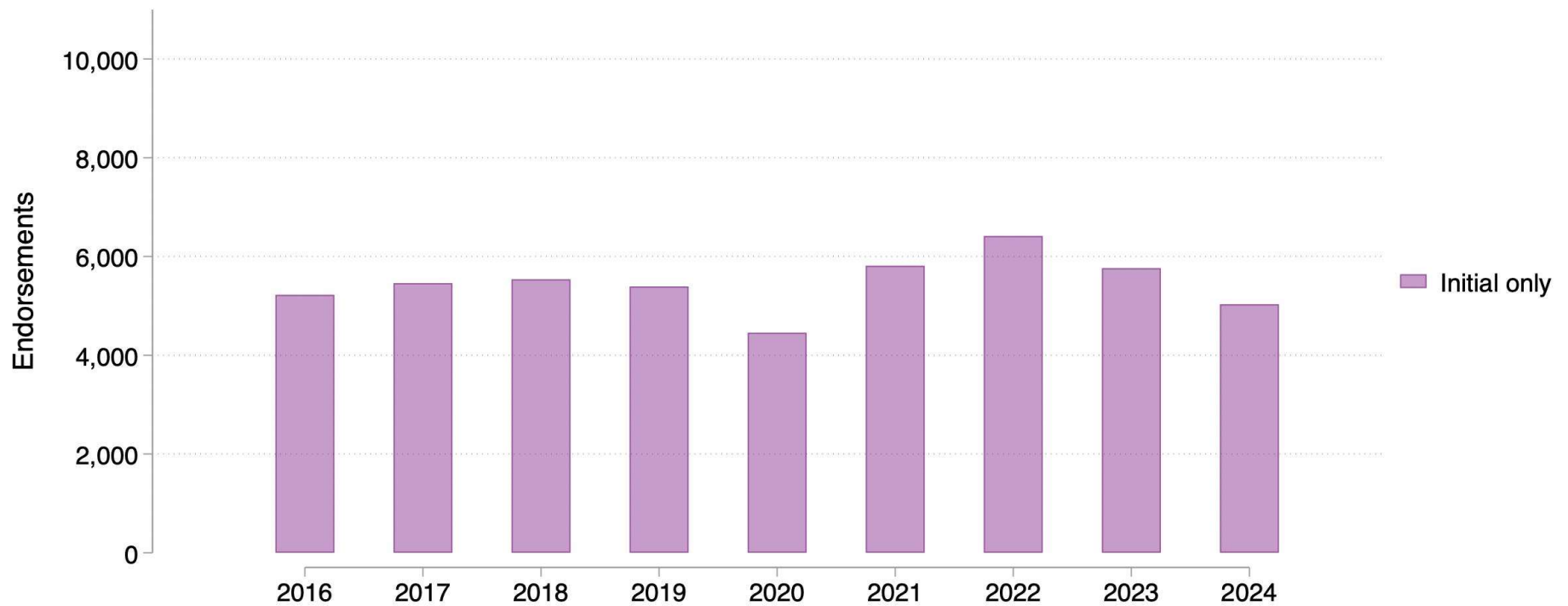
Data and Measures

- It is difficult/impossible to get a national picture of the teacher labor market (Bleiberg & Kraft, 2023), so we focus on a single state: **Washington**
- **Measure of teacher supply:** New teacher credentials and endorsements
- **Measures of teacher demand:**
 - Teacher attrition
 - New teachers hired in state
 - Job postings scraped from school district hiring websites (Goldhaber et al., 2024)

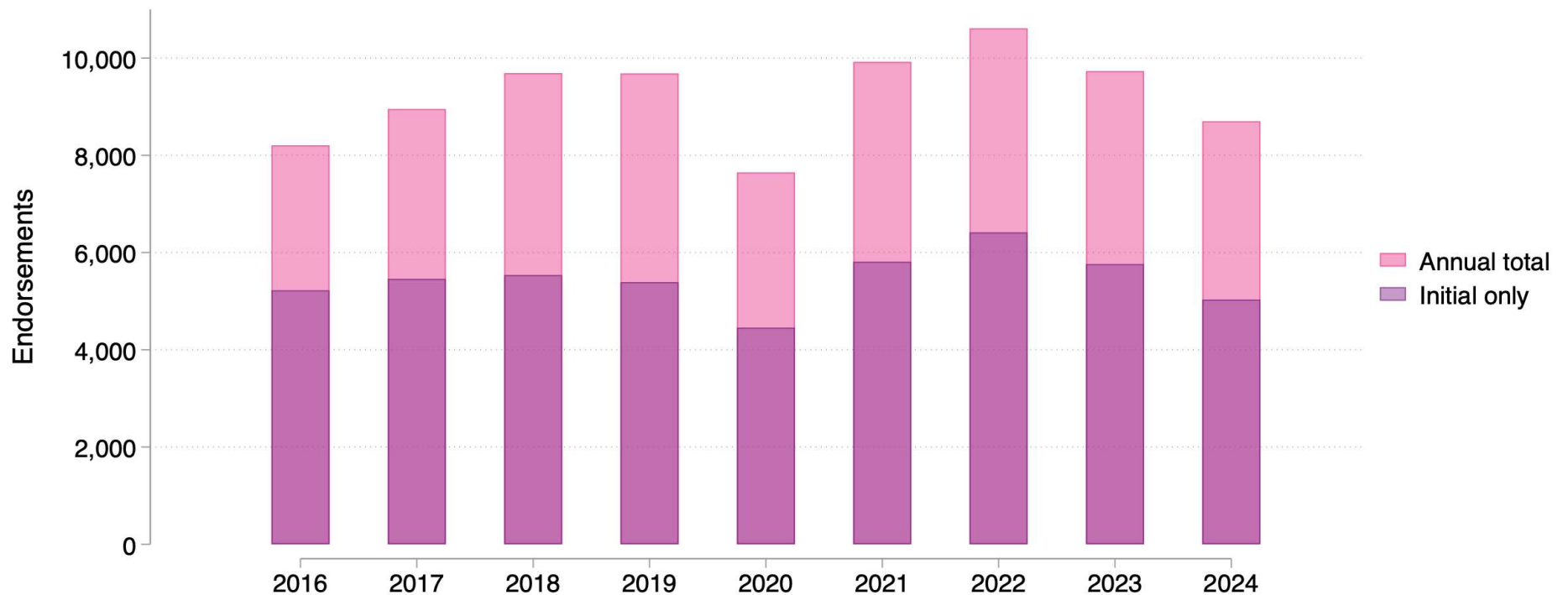
Research Questions

- 1. How have supply and demand for teachers changed relative to prepandemic years?**
2. How does postpandemic demand for teachers differ across specific subject areas and relative to the recent supply of new teachers in each subject area?

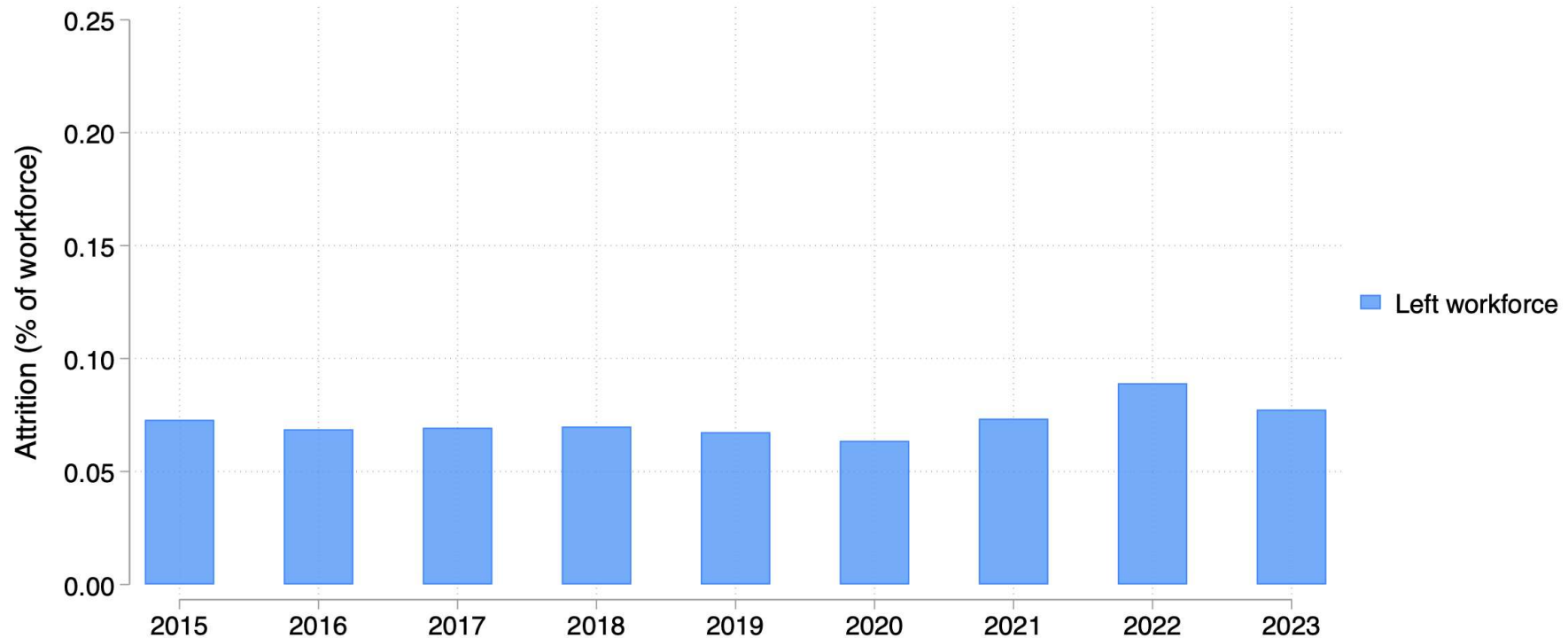
Teacher Supply Over Time in Washington



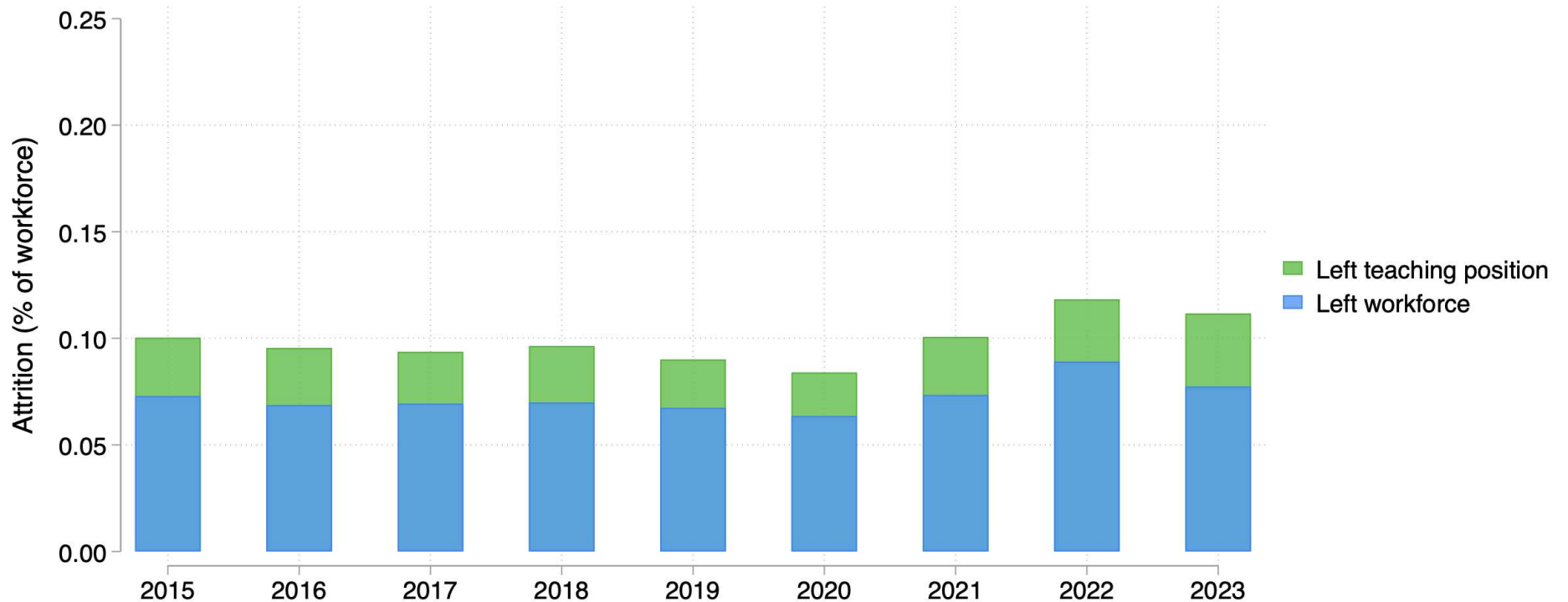
Teacher Supply Over Time in Washington



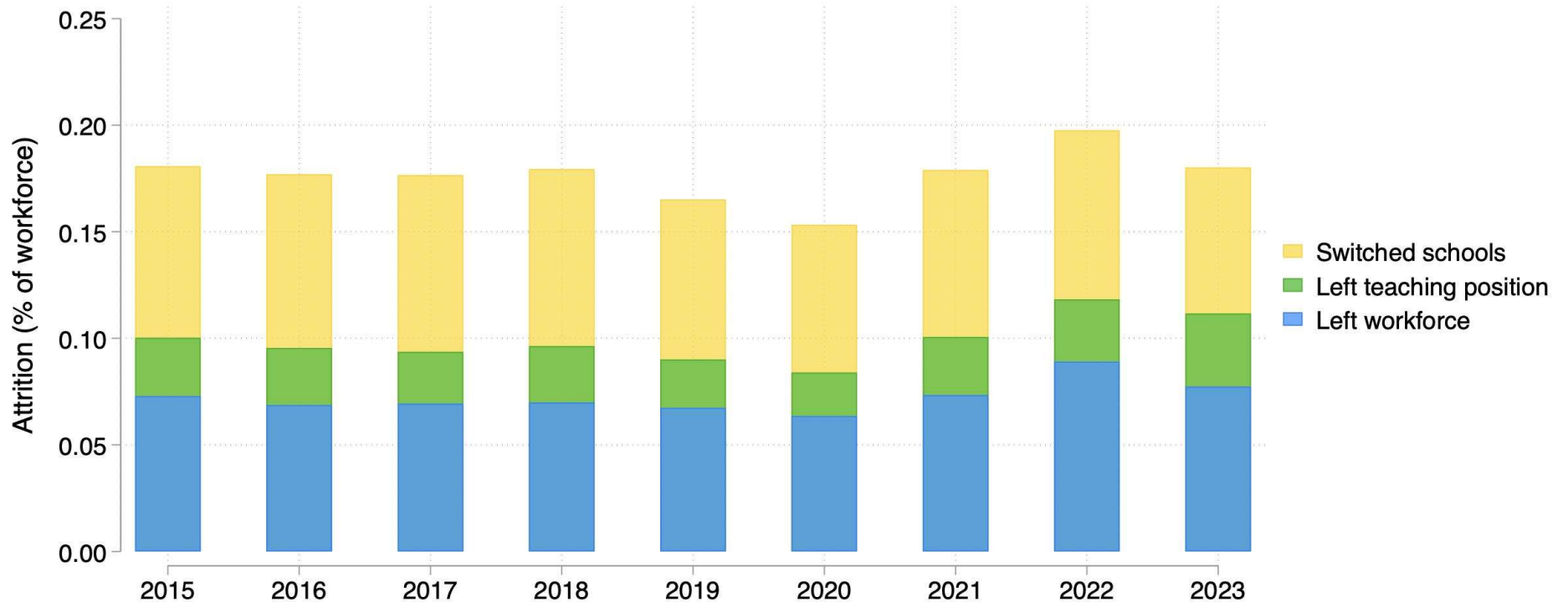
Teacher Attrition Over Time in Washington



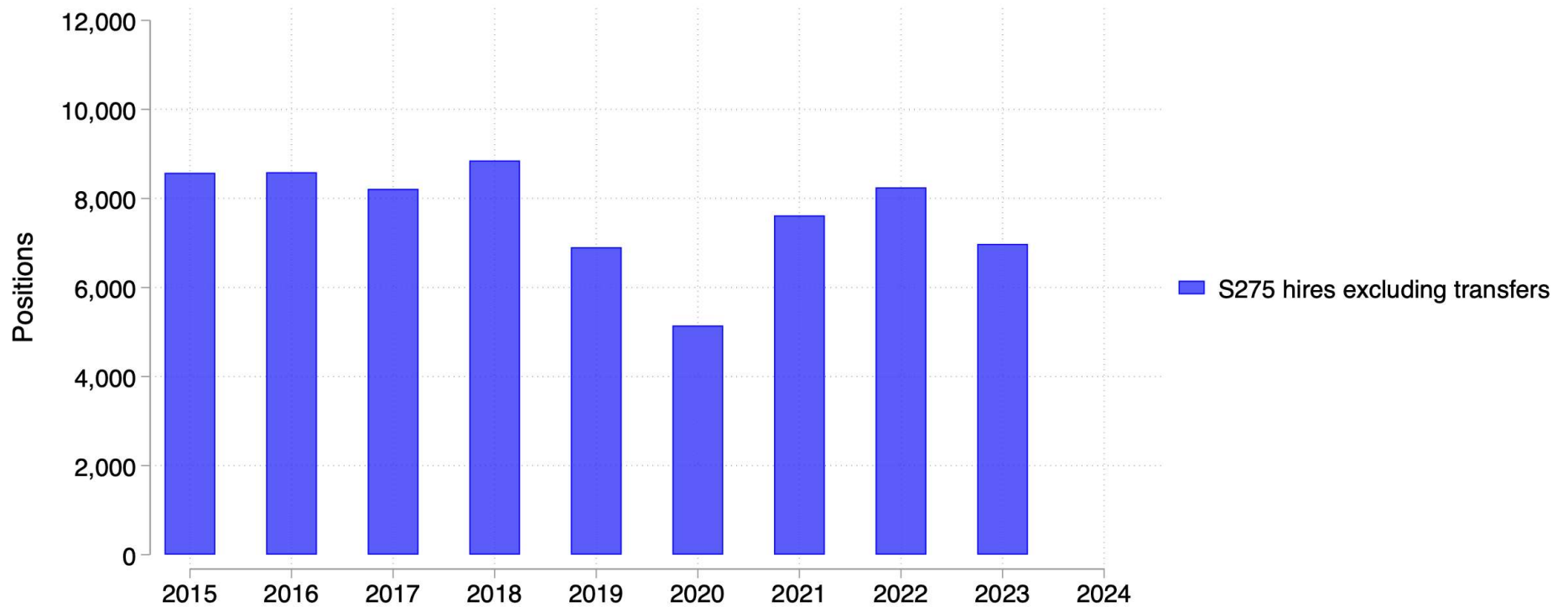
Teacher Attrition Over Time in Washington



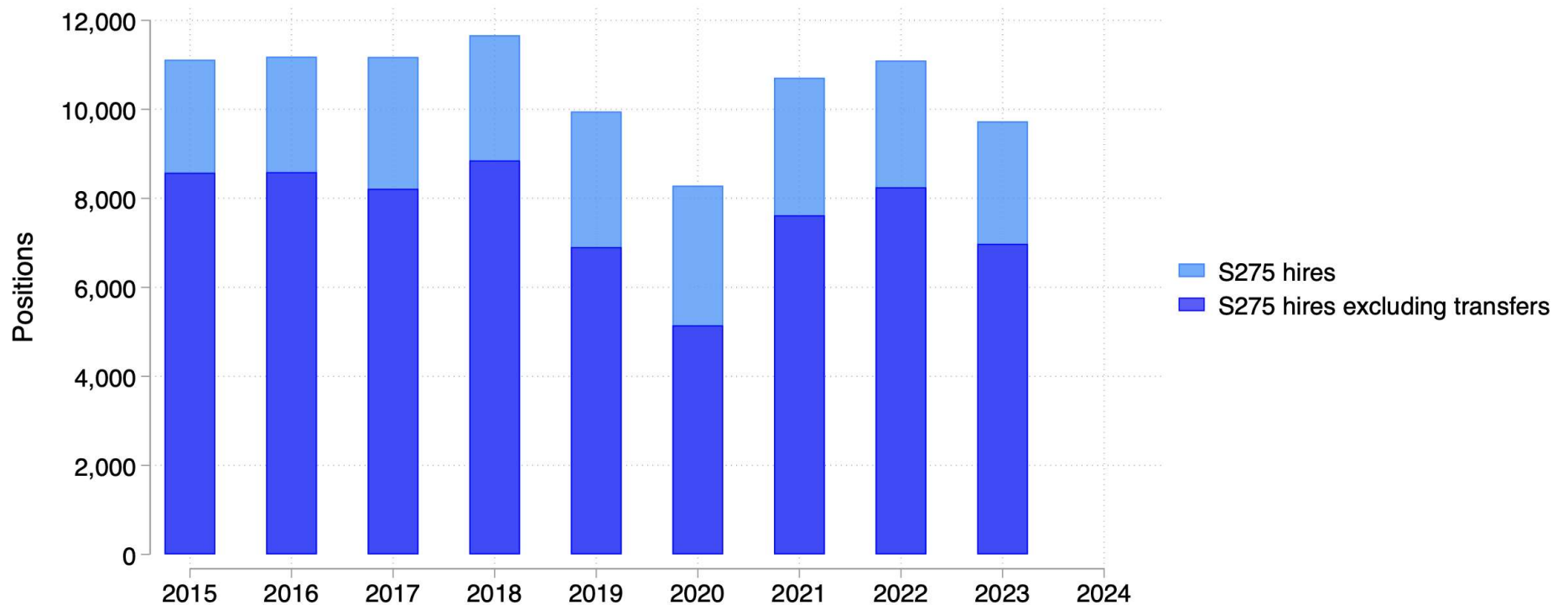
Teacher Attrition Over Time in Washington



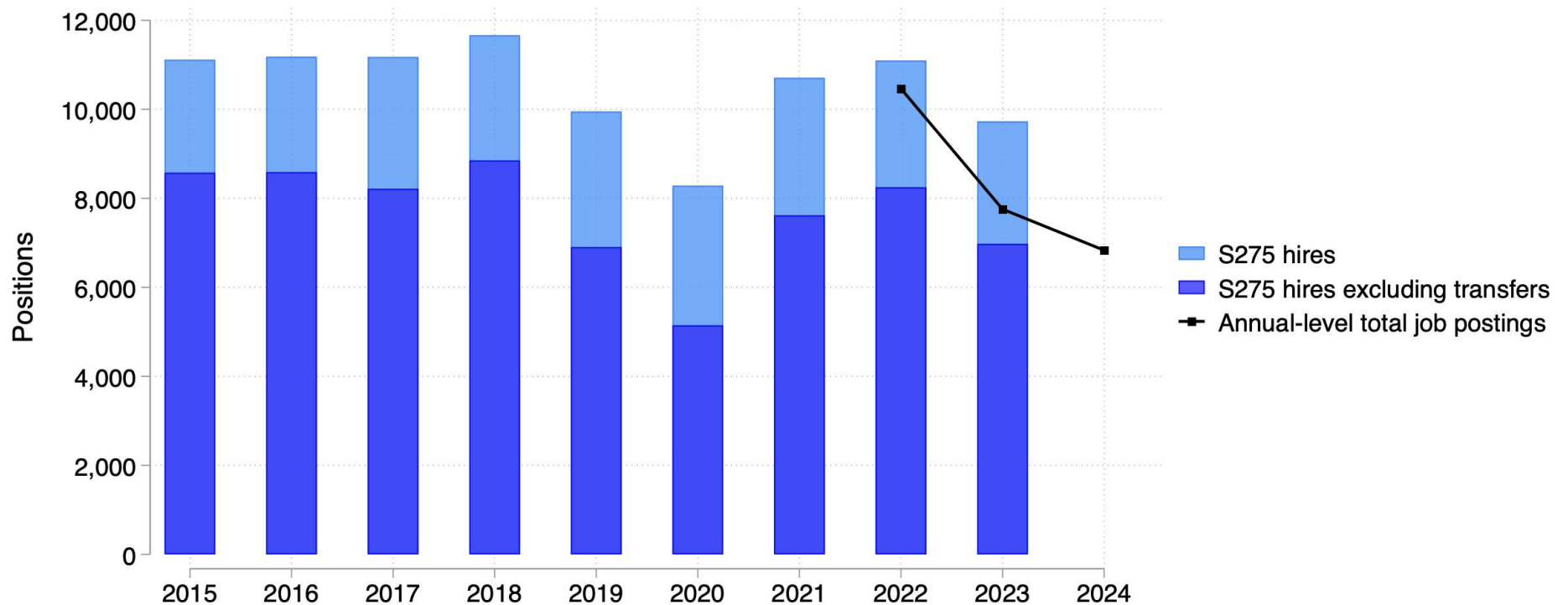
Teacher Hiring and Job Postings Over Time in Washington



Teacher Hiring and Job Postings Over Time in Washington



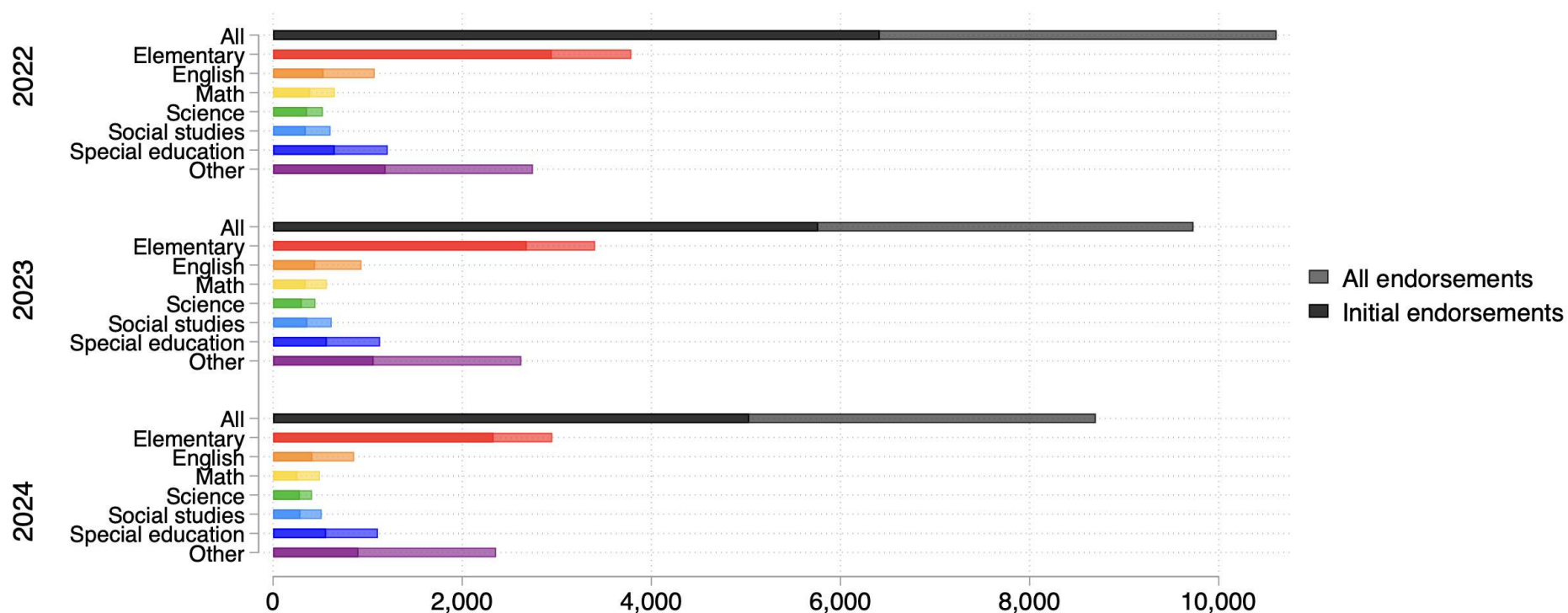
Teacher Hiring and Job Postings Over Time in Washington



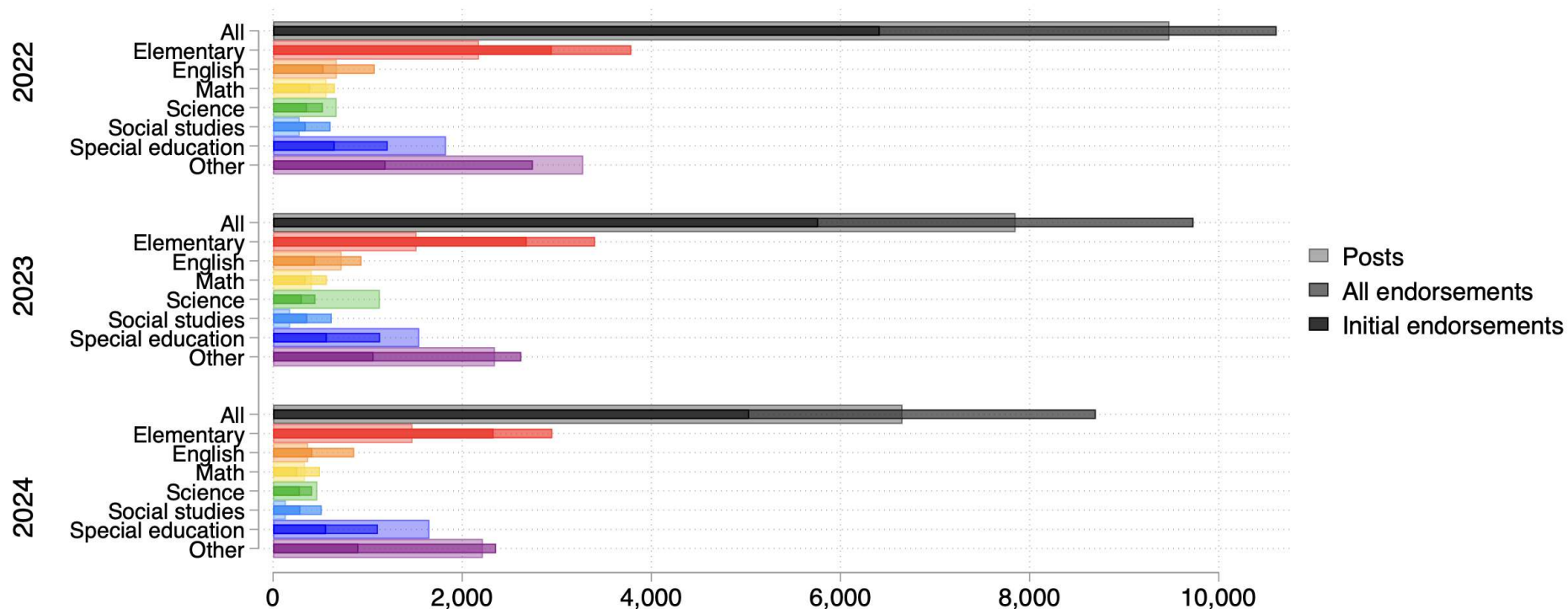
Research Questions

1. How have supply and demand for teachers changed relative to prepandemic years?
- 2. How does postpandemic demand for teachers differ across specific subject areas and relative to the recent supply of new teachers in each subject area?**

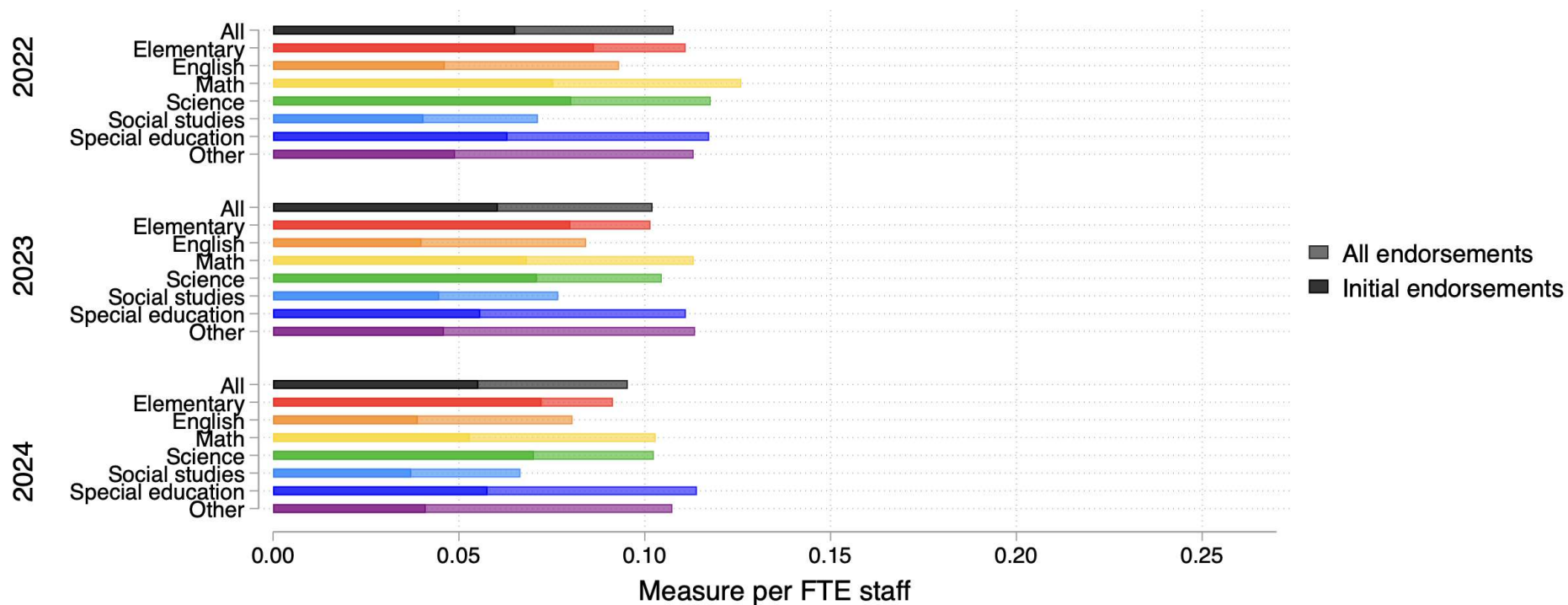
Supply and Demand Overall and by Subject Area in Washington



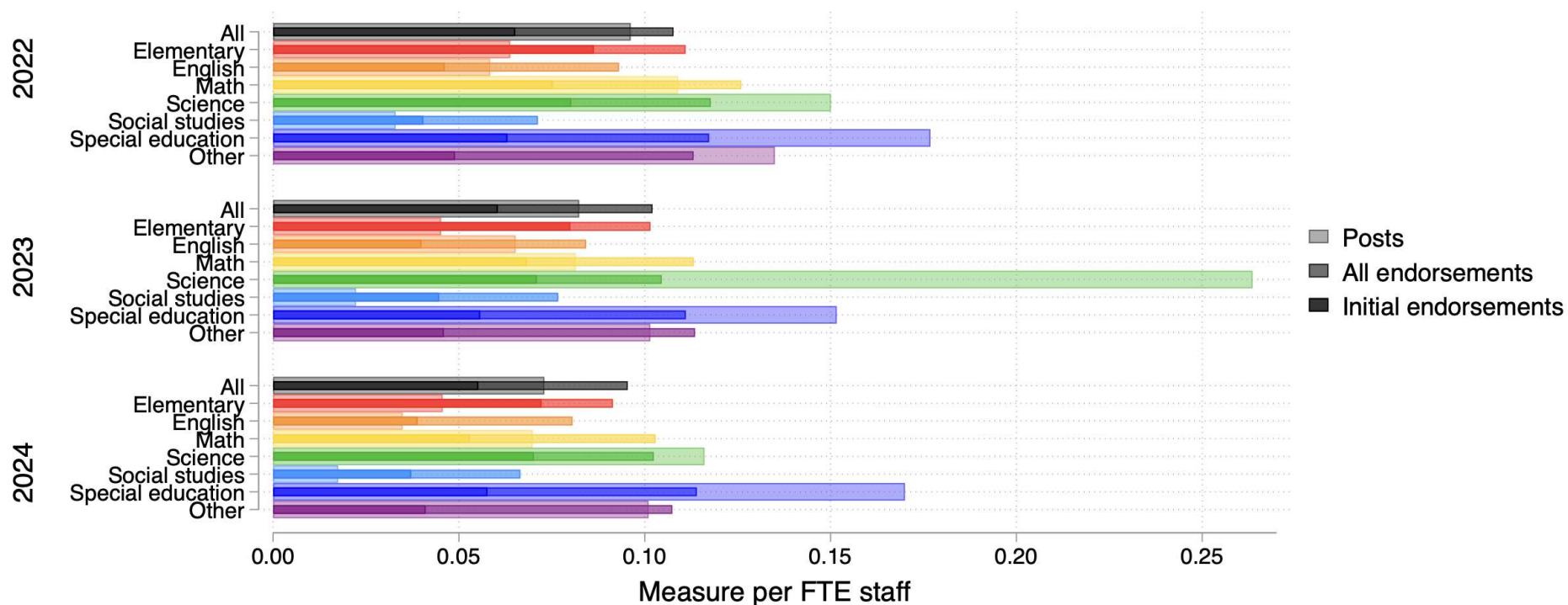
Supply and Demand Overall and by Subject Area in Washington



Supply and Demand Overall and by Subject Area in Washington



Supply and Demand Overall and by Subject Area in Washington



Conclusions

- Teacher demand and supply are **both** down in recent years, and supply slightly exceeds demand in the aggregate
- But demand for special education and science teachers continues to outpace demand for teachers in other subject areas, and supply continues to fall below what is needed to cover demand in these areas
- Teacher recruitment and retention efforts that target these shortage areas (e.g., \$10,000 special education teacher bonus in Hawai'i; Theobald et al., 2025) remain a vital area for policy and future research

The Decline in Teacher Working Conditions During and After the COVID Pandemic

Cory Koedel, University of Missouri/CALDER

Sofia P. Baker, University of Missouri

Motivation

- A lot changed in schools with the COVID pandemic.
- Working conditions are very important to teachers. Teachers often report valuing working conditions more than other aspects of their jobs, including salaries and the socioeconomic status of their students (Horng, 2009; Johnson, Kraft, and Papay, 2012).
- What happened with teacher working conditions during the pandemic?
- What about after the pandemic?

Contribution

- The primary contribution of this project is to bring high-quality data to bear on the question of how teacher working conditions changed during and after the pandemic.
- Our data are from the 5Essentials (5E) Survey in Illinois. Our data panel spans the school years 2016-17 to 2022-23, covering periods before, during, and after the pandemic.
- Our dataset has two key features:
 - The 5E Survey is an incredibly rich survey that captures many dimensions of teacher working conditions.
 - The coverage of schools in the 5E Survey dataset is very high, as is the teacher response rate.
- We are not aware of any datasets that match the depth and coverage of the 5E Survey for measuring teacher working conditions. We believe our data allow us to provide credible evidence on how teacher working conditions have changed in Illinois schools since 2016-17 (hat tip: Paul Bruno).

Preview of Findings

- Teacher working conditions were trending positively going into the pandemic, declined with the onset of the pandemic, and declined at a faster rate in the post-pandemic period.
- Of 20 indicators of teacher working conditions on the 5E Survey, 18 declined from 2019 to 2023.
- The largest declines were in classroom disruptions, the quality of student discussions, expectations for postsecondary education, and teacher innovation (do teachers try hard to improve? Try new ideas?). The trend in teacher safety is also concerning.
- Our study is descriptive and does not isolate the precise cause(s) of the decline in teacher working conditions. I will try not to opine too much.

Data (Survey Content)

- The 5E Survey dataset measures teacher working conditions deeply.
- The 5E Survey asks teachers 105 questions grouped into 20 school-level **Measures** of teacher working conditions.
 - The survey is developed and administered by the Chicago Consortium on School Research (CCSR).
 - The CCSR uses Rasch analysis, which is IRT-based, to improve inference from the survey.
 - All Measures are standardized so a value of 50 indicates the 2013 Illinois average. A 20-point move is equivalent to one standard deviation in the 2013 distribution.
 - Here are the 20 Measures of teacher working conditions (i.e., all of the teacher-directed Measures on the Survey):

Classroom Disruptions	Collaborative Practices	Collective Responsibility	Teacher Influence	Innovation
Instructional Leadership	Program Coherence	Parental Influence	Parental Involvement	Student Responsibility
Professional Development	Reflective Dialogue	School Commitment	Quality of Student Discussion	New Teacher Socialization
Teacher-Parent Trust	Teacher-Principal Trust	Teacher-Teacher Trust	Teacher Safety	Postsecondary Expectations

Data (Survey Coverage)

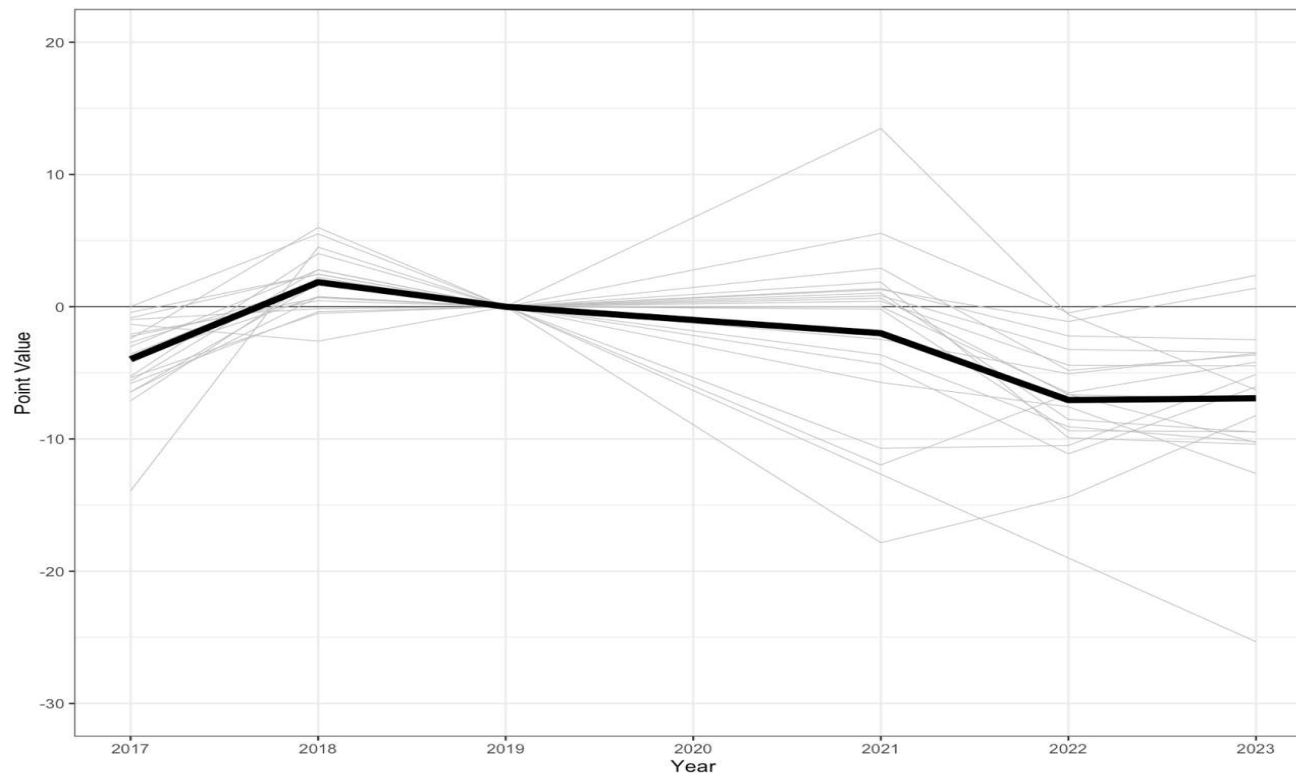
- The coverage of the 5E Survey over Illinois schools and teachers is excellent.
- Just before the 2018-19 school year, Illinois incorporated the 5E Survey into its ESSA plan to measure learning conditions in schools. From this year forward the survey was administered annually in most Illinois schools.
- Conditional on school participation, the teacher response rate has also been very high.
- Prior to 2018-19 the 5E Survey was also administered in the state, but not as universally.
- The 5E Survey was not administered during the 2019-20 school year because the pandemic hit during the survey window in the spring.

Data (Survey Coverage)

Table 1. 5E Survey sample size and coverage by year.

Year (spring)	Schools & Students				Teachers		
	5E Schools	5E Schools as a fraction of all schools listed in the CCD	Number of students in 5E schools	Students in 5E schools as a fraction of total Illinois enrollment in the CCD	Number of teachers in 5E schools*	Teacher survey response rate (Conditional on School Sampling)	Estimated number of teacher respondents (number*rate)
2023	3,524	0.80	1,725,630	0.93	167,527	77.13	129,219
2022	3,393	0.77	1,690,481	0.90	164,118	75.29	123,568
2021	3,436	0.78	1,707,862	0.90	164,095	76.22	125,078
2019	3,511	0.80	1,807,721	0.91	159,232	81.68	130,068
2018	2,029	0.47	1,096,685	0.55	84,013	86.77	72,896
2017	2,175	0.52	1,105,975	0.55	80,140	85.25	68,323

Trends in Teacher Working Conditions (Visualization)



Trends in Teacher Working Conditions

Table 3. Changes in teacher working conditions over various time periods.

	During & After the Pandemic	2-Year Windows (Spring to Spring)		
	2019-2023	2017-2019	2019-2021	2021-2023
Level of Classroom Disruptions	-25.3*	13.9*		
Collaborative Practices	-5.1*	5.6*	-10.7*	5.6*
Collective Responsibility	-9.5*	2.7*	1.0*	-10.5*
Teacher Influence	-2.5*	2.1*	1.4*	-3.9*
Innovation	-10.4*	0.4	1.9*	-12.3*
Instructional Leadership	-4.2*	3.9*	0	-4.2*
Program Coherence	-3.6*	5.8*	2.9*	-6.6*
Parent Influence on Decision Making in Schools	2.4*	0.8	5.6*	-3.2*
Parent Involvement in School	-3.5*	7.1*	-2.5*	-1.1*
Student Responsibility	-10.3*	0	-12.0*	1.7*
Quality Professional Development	-6.1*	6.4*	-4.3*	-1.7*
Reflective Dialogue	-8.2*	3.6*	-17.8*	9.6*
School Commitment	-9.5*	2.2*	-0.2	-9.3*
Quality of Student Discussions	-12.6*	2.4*	-5.7*	-6.9*
Socialization of New Teachers	1.4*	1.3	1.3*	0.1
Teacher-Parent Trust	-6.9*	5.3*	0.6	-7.6*
Teacher-Principal Trust	-3.5*	1.0	0.8	-4.3*
Teacher-Teacher Trust	-4.5*	5.4*	0.4	-4.9*
Teacher Safety	-6.3*	3.0*	13.5*	-19.8*
Expectations for Postsecondary Education	-10.2*	6.5*	-3.6*	-6.6*
Simple Average	-6.9	4.0	-1.4	-4.5

Trend Heterogeneity

- Although schools serving high-SES students have much better working conditions in levels than schools serving low-SES students, the trends in teacher working conditions from 2019-2023 are similar for both groups.
- The declines in working conditions have been larger in Chicago Public Schools than in other Illinois districts, but the general pattern of estimates is similar.
 - Most notably, the decline from 2021 to 2023 was much larger than from 2019 to 2021 both in CPS and outside.
- Illinois was in the top quartile of time spent in online learning during the 2020-21 school year (Goldhaber et al., 2023).
 - Schools in districts that spent more time in the online mode of instruction had significantly larger declines in teacher working conditions compared to schools in districts that spent more time in person.
 - This is an interesting result, with the caveat that the root cause it not entirely clear.

Conclusion

- I hope I have convinced you that the 5E Survey dataset is an exceptionally high-quality dataset for documenting changes in teacher working conditions over time.
 - The depth of the 5E Survey gives us confidence that the trends we have identified reflect teachers' true perceptions of their working conditions—i.e., we are not depending on a small number of survey items to draw broad conclusions. The high coverage of the survey, and high teacher response rate, limit concerns about biased inference due to sample selection.
- If so, there are not too many ways for our simple analysis to be misleading.
 - We check some basic items, such as looking for a dramatic shift in the composition of the workforce, and nothing reveals itself.
 - The trends we document are similar across subsamples of the data.

Conclusion

- Two limitations of our study are:
 - Our sample is limited to teachers in Illinois.
 - We have not identified the causal mechanisms driving the changes to teacher working conditions. This is an area ripe for future research. However, disentangling the many potential factors that could have contributed to these trends during the tumultuous period in public schools since 2019 will be difficult, and it seems unlikely that researchers will have timely answers.
 - » Policymakers and education officials may need to make their best guesses.

What Do We Know about Academic Recovery Interventions?

Emily Morton, CALDER/AIR
A Cast of Thousands

Slides for this paper will be added when
results are public and ready to share

ESSER II & II and Student Achievement: Assessing the Impacts of the Largest One-Time Federal Investment in K12 Schools

Dan Goldhaber, CALDER/AIR & University of Washington

Grace Falken, University of Washington

What This is About

The New York Times

***Schools Got a Record \$190 Billion in
Pandemic Aid. Did It Work?***



By Sarah Mervosh

EducationWeek®

**What the Historic Infusion of
Federal Pandemic Aid Did for
Schools**



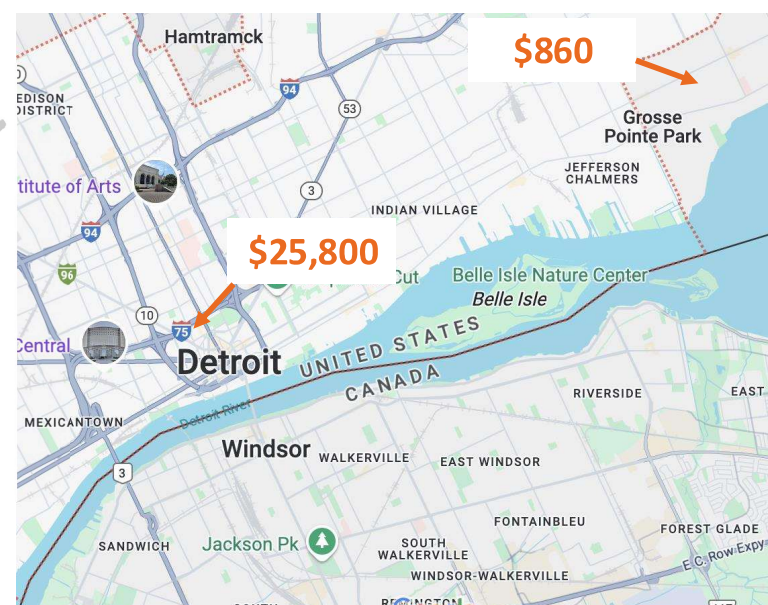
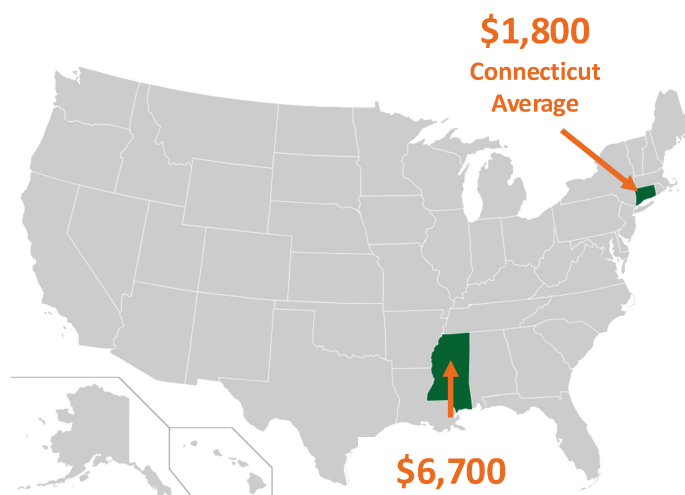
By Sarah D. Sparks

COVID Relief Was Massive, Targeted by Need

ESSER is more than triple the federal funding allocated to districts in 2019-20, and four times the extra funding sent to districts under ARRA of 2009

ESSER II (Winter 2020) and **ESSER III** (Spring 2021) based on Title I for 2020-2021, allocated based on 2018 SAIPE

20% of ESSER III designated to address pandemic learning loss



You Probably Remember This: 4th Grade NAEP Math Scores

FIGURE | Trend in fourth-grade NAEP mathematics average and selected percentile scores



Between 2019 and 2022, 4th grade math scores dropped $\sim .15$ SD

Between 2022 to 2024, 4th grade math scores recovered by $\sim .03$ SD

What We're Doing

- Looking at the impact of ESSER spending in 2022-23 on test scores in spring 2023 in 28 states

ESSER spending → Student Achievement

- But! There are good reasons to worry that estimates of ESSER effects on achievement will be biased (i.e., misleading)
 - Districts with different ESSER allocations or spending rates could have distinct unobserved characteristics that affect both funding and achievement
 - » Higher-poverty districts tend to be in communities that faced disproportionately negative COVID impacts (e.g., health and employment outcomes)
 - » District that spent ESSER funding more quickly might have higher organizational capacity

Wading a *Little* into the Econometric Details

- We instrument for ESSER spending per pupil with Title I formula-eligible percent (FEP) from 2018
 - Identifying assumption: conditional on district characteristics including current poverty rate, the variation in FEP is plausibly exogenous
- Three sources of identifying variation
 - Time lag in FEP reporting (2018 values determined ESSER \$)
 - Differences between geo area poverty vs enrolled student poverty
 - Measurement error (FEP is an estimate, FRPL is reported by schools)

How to Think About ESSER and the Role of States

Do we trust estimates *without* state fixed effects (relying on within and between variation in ESSER spending) or those *with* state fixed effects (estimates based only on within state variation in ESSER)?

No State Fixed Effects

- A small share of the variation in ESSER funding is associated with states (e.g., small state minimums)
- 10% of ESSER funding went to states so some of the “ESSER effect” may be driven by state-level ESSER spending decisions
 - States were prohibited from mandating how districts spent ESSER funds (U.S. Department of Education, 2022), but have influenced decisions, timelines, and recovery through their approval of district spending plans (Lieberman, 2022)

Include State Fixed Effects

- We do not want to conflate state-level differences in policy or context with the influence of ESSER spending on student outcomes

We Use Publicly-Available Data

- Test scores—Stanford Data Archive (SEDA)
 - 30 states measured in 2018-19, 2021-22, and 2022-23
 - Grades 3-8 math and ELA
- District characteristics—Common Core of Data (CCD)
 - Student demographics, pre-pandemic finance measures, urbanicity
- ESSER Spending—Collected by us, directly from states: 28 respondents
 - If you know anyone in Alabama and/or Oklahoma’s Departments of Education, let us know!

What Did We Find?

- ESSER funding helped address pandemic-related academic challenges
- Results differed by subject:

Table 3. Estimated impacts of ESSER II and III spending on district-level student achievement

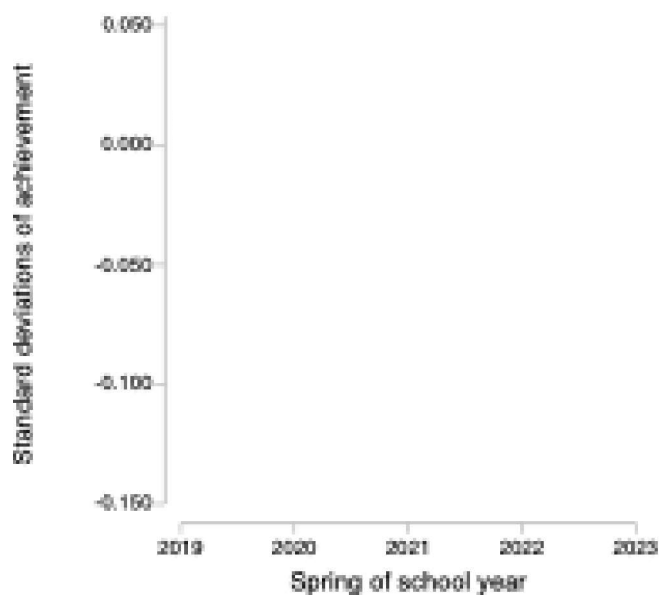
	(1)	(2)	(3)	(4)
<i>Panel A. Math achievement</i>				
ESSER per pupil spending	0.005***	0.006**	0.019***	0.007*
2022-23 (\$1,000s)	(0.002)	(0.003)	(0.005)	(0.004)
<i>n</i>	4,989	4,989	4,989	4,989
<i>Panel B. English language arts achievement</i>				
ESSER per pupil spending	0.021***	0.006***	0.019***	0.002
2022-23 (\$1,000s)	(0.004)	(0.002)	(0.004)	(0.003)
<i>n</i>	4,824	4,824	4,824	4,824
Estimator	OLS	OLS	2SLS	2SLS
State fixed effects		X		X

Sensitivity to Additional (State-Level) Controls

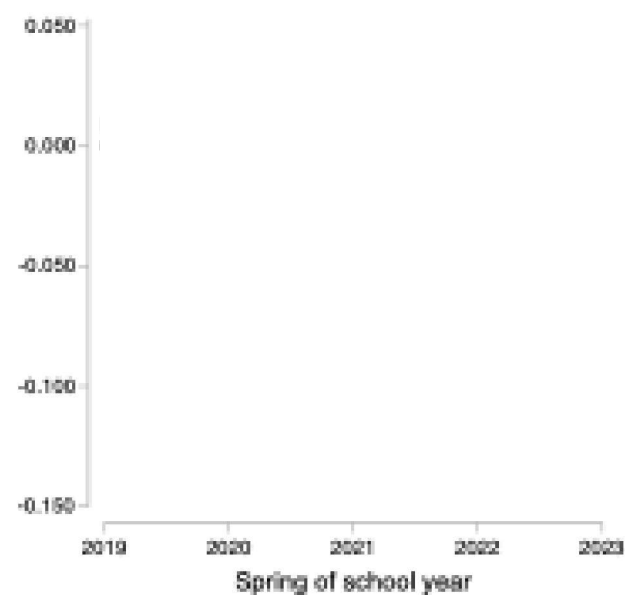
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A. Math 2023 achievement											
ESSER per pupil spending 2022-23 (\$1,000s)	0.021*** (0.005)	0.020*** (0.006)	0.019*** (0.005)	0.016*** (0.005)	0.017*** (0.005)	0.019*** (0.005)	0.019*** (0.005)	0.018*** (0.005)	0.017*** (0.005)	0.018*** (0.005)	0.008* (0.004)
n	4,989	4,989	4,989	4,989	4,989	4,989	4,989	4,989	4,989	4,989	4,989
Panel B. English language arts 2023 achievement											
ESSER per pupil spending 2022-23 (\$1,000s)	0.018*** (0.005)	0.021*** (0.004)	0.020*** (0.004)	0.017*** (0.005)	0.013*** (0.004)	0.021*** (0.004)	0.018*** (0.005)	0.018*** (0.004)	0.024*** (0.004)	0.020*** (0.004)	0.003 (0.003)
n	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824	4,824
District % instruction remote and hybrid	X										X
State youth access to computer and internet		X									X
State COVID cases and deaths			X								X
State democratic vote share 2020				X							X
State unemployment rate					X						X
State total K12 education revenue 2019-20						X					X
State indicators for Right to Work							X				X
State additional Title I formula inputs								X			X
CRPE student performance saliency									X		X
State ESSER % spent prior to and during 2022-23										X	X

What Does that Look Like?

Panel A. Math

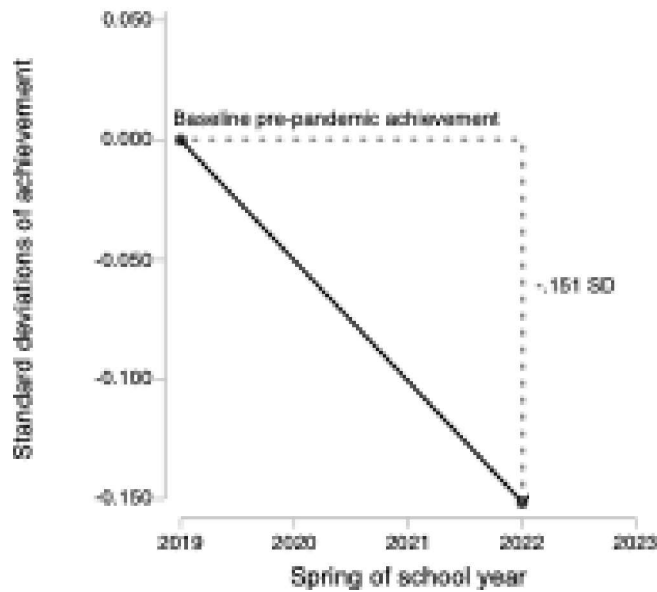


Panel B. ELA

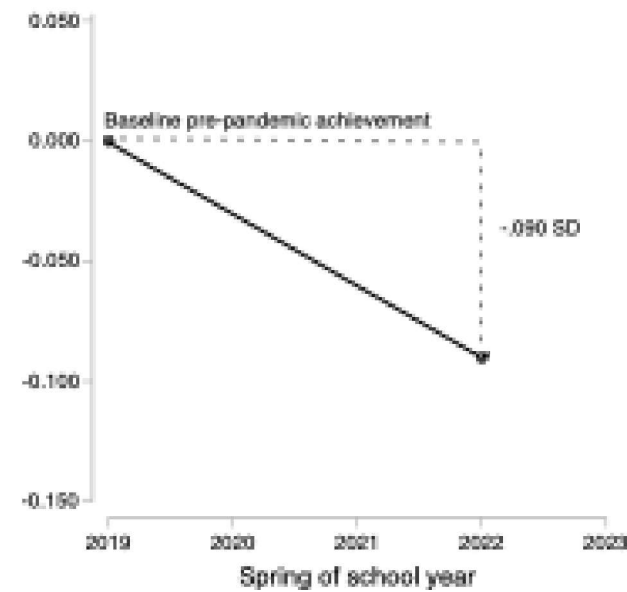


Historic Losses in Math and ELA Between 2019 and 2022

Panel A. Math

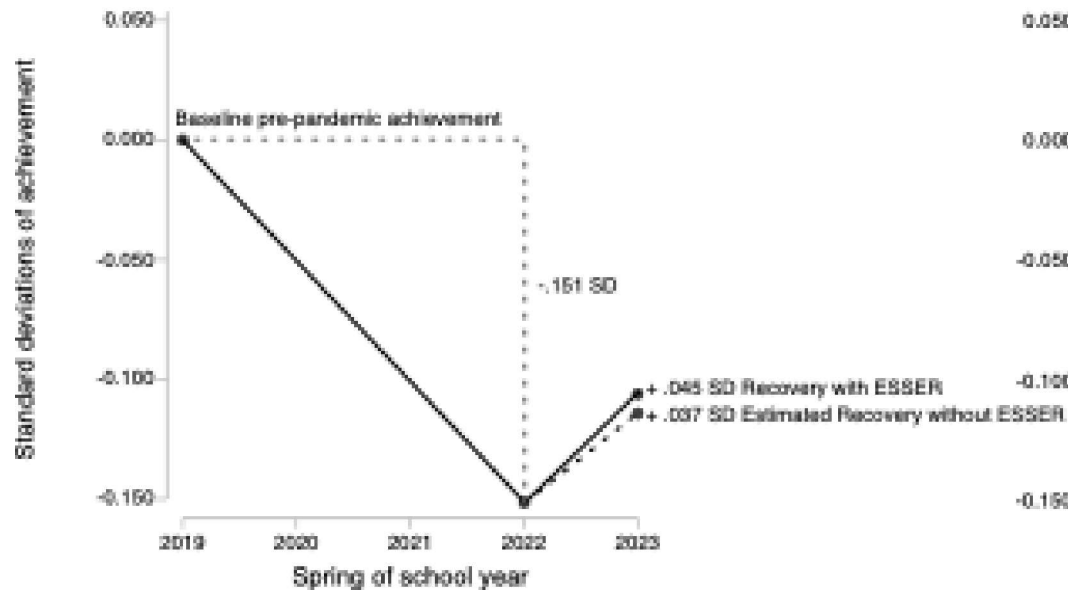


Panel B. ELA

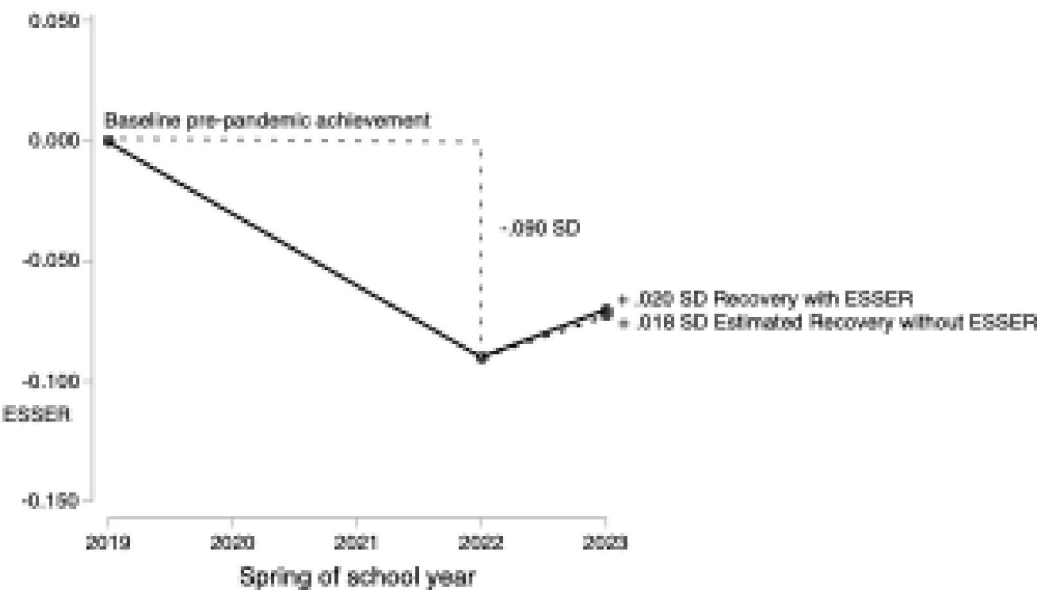


Estimated Recovery With and Without ESSER

Panel A. Math

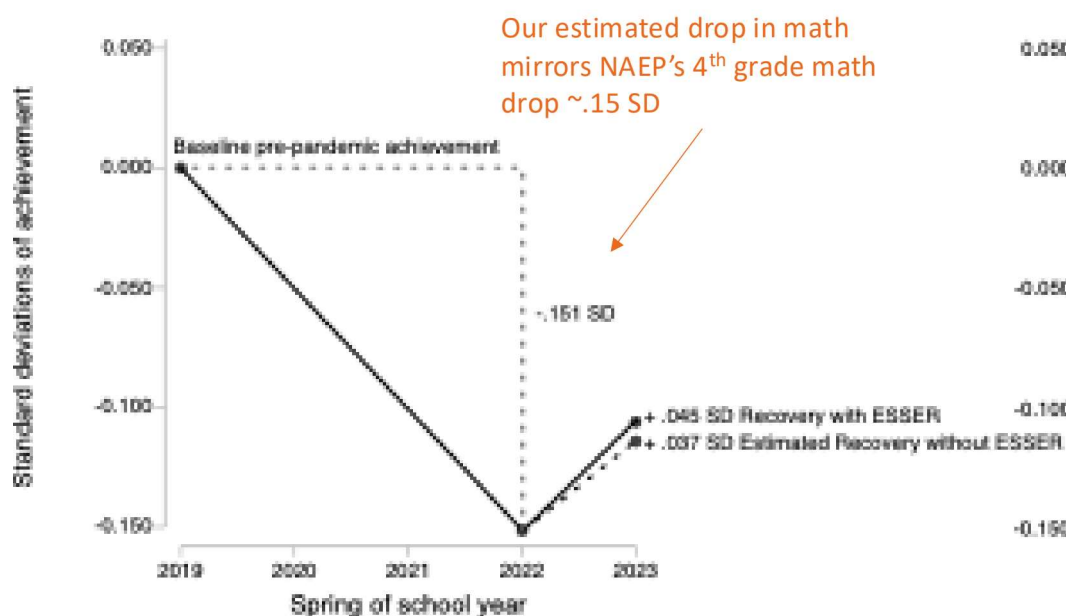


Panel B. ELA

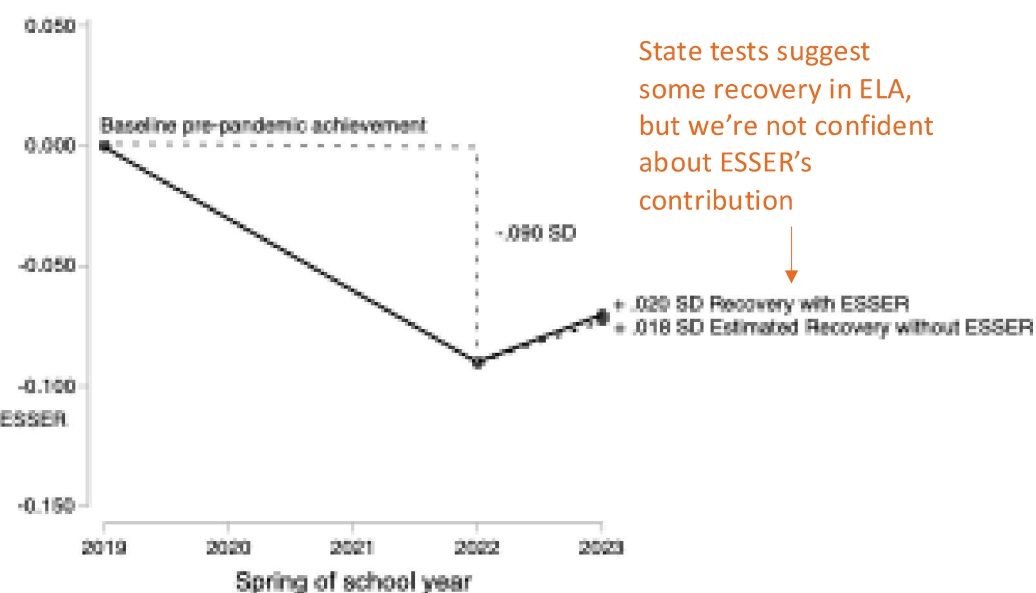


Estimated Recovery With and Without ESSER

Panel A. Math



Panel B. ELA



A Rorschach Test

- For math, in-line with findings from a premier meta-analysis on the impacts of funding on achievement outcomes (Jackson and Mackevicius, 2024). ELA results clearly disappointing.
- At the scale we're talking about (~28M students), a .007 SD impact on math for a \$1,000 investment is large (Kraft, 2022).
- Assuming the return is the same, estimates suggest \$15,380 more per pupil needed for full math recovery.
- Even supporters of the funding think funding could have been used better, more accountability. *How* schools use funding matters (Handel & Hanushek, 2023; McGee, 2023).
- Interpreting state fixed effects—importance of state role.



Session II: Predicting Students' Post-secondary Success

Paper Presentations

- **The Geography of Educational Opportunity: Teacher Quality.** Steven Rivkin, University of Illinois Chicago, CALDER
- **The Expansion of Multiple-Measure Placement Policies During COVID-19: A Case Study of Kentucky Community Colleges.** Zeyu Xu, CALDER/AIR
- **Disrupted Trajectories: College Enrollment and Employment Among High School Graduates Through the COVID-19 Pandemic.** Steven Hemelt, University of North Carolina at Chapel Hill, CALDER

Reactor

- **Chris Mullin,** Lumina Foundation

The Geography of Educational Opportunity: Teacher Quality

Benjamin Feigenberg, University of Illinois Chicago

Eric Hanushek, Stanford University/CALDER

Samuel Mosley, University of Illinois Chicago

Steven Rivkin, University of Illinois Chicago/CALDER

Results and opinions do not reflect the views of any Texas state agency, the state of Texas, or the Texas Schools Project.

Very preliminary, please do not cite without authors' permission.

An Investigation of the Geography of Education Opportunity

- How much do schools contribute to the geographic variation in social mobility and educational, social and economic outcomes
 - In two papers, Chetty and coauthors find differences in schooling outcomes are associated with differences in social mobility
- We use administrative data and a rigorous empirical strategy to ask
 - What are the contributions of differences in educator effectiveness including teachers and principals?
 - Through what skill-development channels do schools and educators affect different outcomes?

Research Questions in the Study of Elementary and Middle School Teacher Quality Effects

- Are teacher VA to student cognitive and socio-emotional skills related to longer-term academic, economic and social outcomes?
- Is there substantial geographic variation?
- Preliminary draft: await data on arrests; have not estimated contributions to geographic variation in outcomes

It is Challenging to Measure School-average Teacher Effects on Skill Acquisition

- Nonrandom sorting of students into neighborhoods, schools and classrooms.
- Nonrandom sorting of teachers among schools
- Confounding effects of school and district administrators and policies
- Confounding effects of neighborhood factors including
 - Local economic conditions
 - Criminal activity

Use Teacher Movers to Account for School and Neighborhood Factors and Compare Teacher Quality

- Mr. Hanushek teaches 4th grade math in School A in 2017 and School B in 2018
 - In School A Mr. Hanushek's class has average test score growth of .2 SD
 - In School B Mr. Hanushek's class has average test score growth of .1 SD
- Ignoring test error and assuming Mr. Hanushek's true effectiveness does not change over time
 - The difference between average growth in A and B reflects differences in school and neighborhood factors in each year

Creating a Relative Index of Average Teacher Effectiveness at Raising Cognitive Skills

- Consider the following:
 - average test score growth for all teachers in School A -0.2 SD >
 - average test score growth for all teachers in School B -0.1 SD
- then average teacher quality is higher in School A
- Let average test score growth in School A = 0.25 SD and average test score growth in School B = 0.09 SD
- The difference in average teacher quality measured relative to the moving teacher equals
 - $(0.25 - 0.2) - (0.09 - 0.10) = 0.06$ SD

Three Dimensions of Teacher Effectiveness

- Math and reading achievement are proxies for cognitive skills
- Absences and out of school suspensions serve as proxies for potentially different dimensions of socioemotional skills
 - Out of school suspensions have been shown to be strongly related to future arrests (Deming 2011)
 - Related work find they have much stronger effects than receipt of a disciplinary infraction
 - They potentially capture more serious disciplinary infractions, but the penalty itself may elevate the probability of future arrest

Estimation of School Average Teacher Value Added to Each Skill Proxy

- Estimate teacher effects (VA) on each skill each year by controlling for prior skills and demographic characteristics
- Average the VA estimates for each of the skill proxies over all teachers in the school in each year
- Use Bayesian methods to shrink estimates toward the grand mean and lessen attenuation bias due to estimation error

Producing the School Average Teacher Quality Index

- Subtract VA for teachers who switch schools from average value added separately for each skill
- Create the teacher quality index from comparisons across teachers who change schools
 - The index is fixed over time
- Take the average index for grades 4-8 for students who change schools

Key Assumption: Differences Between Mr. Hanushek's Performance in Schools A and B Reflects Only Factors that Affect all Classrooms in a School

- What if Mr. Hanushek is assigned relatively better students in School A in ways that the controls do not address? Problem!
- What if Mr. Hanushek improves as a teacher? Problem!
- We adjust Mr. Hanushek's growth measures (VAs) prior to differencing
 - control for having little or no experience, moving prior to the school year, teaching in a different grade, and for the presence of a student classified as emotionally disturbed (in progress)

Merge Information from Four Administrative Data Sets by Unique Student IDs

- Primary and secondary school data come from the TEA
- Higher education data come from the THECB and National Student Clearinghouse for those attending college outside of Texas (results today do not include NSC data)
- Employment data come from Texas Workforce Commission
- CCH data on arrests, court dispositions and incarceration come from the Texas Department of Public Safety and will be added in the future

Data

- Unique IDs allow students and educators to be followed over time, even if they change schools, as long as they remain in a Texas public school
- Hundreds of thousands of students
- Thousands of elementary and middle schools
- Student-teacher matches are available only for 2012 forward
- Current results focus on the 2019 high school graduation cohort
 - We will expand the sample in the future

Table 1. Estimated Standard Deviations of Teacher Effectiveness Relative to Other Schools in the Largest Connected Network (89% of schools)

Aggregation level	Teacher	School	School	School	Student average over grades 4-8
Control for school and neighborhood influences	No	No	Yes	Yes	Yes
Shrunk	No	No	No	Yes	Yes
Math VA	0.283	0.134	0.183	0.071	0.045
Reading VA	0.222	0.119	0.140	0.052	0.032
Absence VA	2.447	2.376	1.332	0.439	0.296
Suspension VA	0.093	0.080	0.051	0.018	0.012

Step 2: Regress Each Outcome on the Value Added Measures and Controls

- College attendance and college persistence into the second year of college
- Strong engagement with employment and school
- Attend college or work in 6 of 8 quarters
- Outcomes are measured in the two-year period following expected high school graduation date given year in 8th grade

Table 3. Estimated Effects of School Average Teacher Effectiveness at Raising Cognitive and Non-cognitive Skills Aggregated over Grades 4 to 8 on College Attendance, College Persistence, and the Probability of Engaging in the Productive Activities of School or Work after High School

	College attendance	College persistence	Engage in a productive activity
Math VA	0.131 (0.042)	0.183 (0.035)	0.101 (0.038)
Reading VA	0.047 (0.060)	0.086 (0.052)	0.041 (0.056)
Absence VA	-0.022 (0.007)	-0.006 (0.006)	-0.008 (0.006)
Suspension VA	-0.624 (0.162)	-0.488 (0.152)	-0.379 (0.152)

Table 4. Simulated Effects of One Standard Deviation Increases in School Average Teacher Effectiveness at Raising Cognitive and Non-cognitive Skills on College Attendance, College Persistence, and the Probability of Engaging in the Productive Activities of School or Work after High School

	College attendance	College persistence	Engage in a productive activity
Math VA	0.009	0.013	0.007
Reading VA	0.002	0.004	0.002
Absence VA	-0.010	-0.003	-0.004
Suspension VA	-0.011	-0.009	-0.007

Sensitivity Analysis and Future Work

- Assuming students remain in the school attended in 4th grade does not alter the results
 - This lessens concerns that purposeful student school changes introduce bias
- We will explore sensitivity to additional classroom controls in the adjustment of the mover VAs
- We will also investigate effects on the probability of engagement with the criminal justice system
- Finally, we will leverage school openings and closings to generate differences in school attended that are not driven by family choices

Effects of Geographic Differences in School Quality

- We focus on variation across schools in this draft. The magnitudes are meaningful but modest, and variation across districts and commuting zones will be even smaller
- We do not capture the impacts of teacher quality differences
 - NOT captured by our three skill proxies
 - In grades k to 3, though we would expect variation in school average teacher quality to be similar in the early and later elementary school grades

The Expansion of Multiple-Measure Placement Policies During COVID-19

A Case Study of the Kentucky Community and Technical College System (KCTCS)

Zeyu Xu | Ben Backes

American Institutes for Research

We thank the National Science Foundation and the W. T. Grant Foundation for financial support, the Kentucky Center for Statistics for providing the requisite data, and the Kentucky Community and Technical College System as our research partner. We appreciate the research assistance provided by Rachel Dao. All errors and opinions are our own.

Assessment and placement policy waiver

About half of incoming KCTCS students are deemed not ready for college-level instruction. Those students are required to receive developmental education (DE) support.

Test-based assessment

- The Kentucky Community and Technical College System (KCTCS) has been primarily using the **ACT** (mandatory for HS juniors).
- Other tests include Accuplacer, ALEKS, EdReady, KYOTE, SAT, etc.



Multiple measure assessment (MMA)

- In response to COVID-19, KCTCS added **HS GPA** to a list of measures that could be used for placement decisions, starting in 2020-21.

Accurate course placement is a challenge to get right.

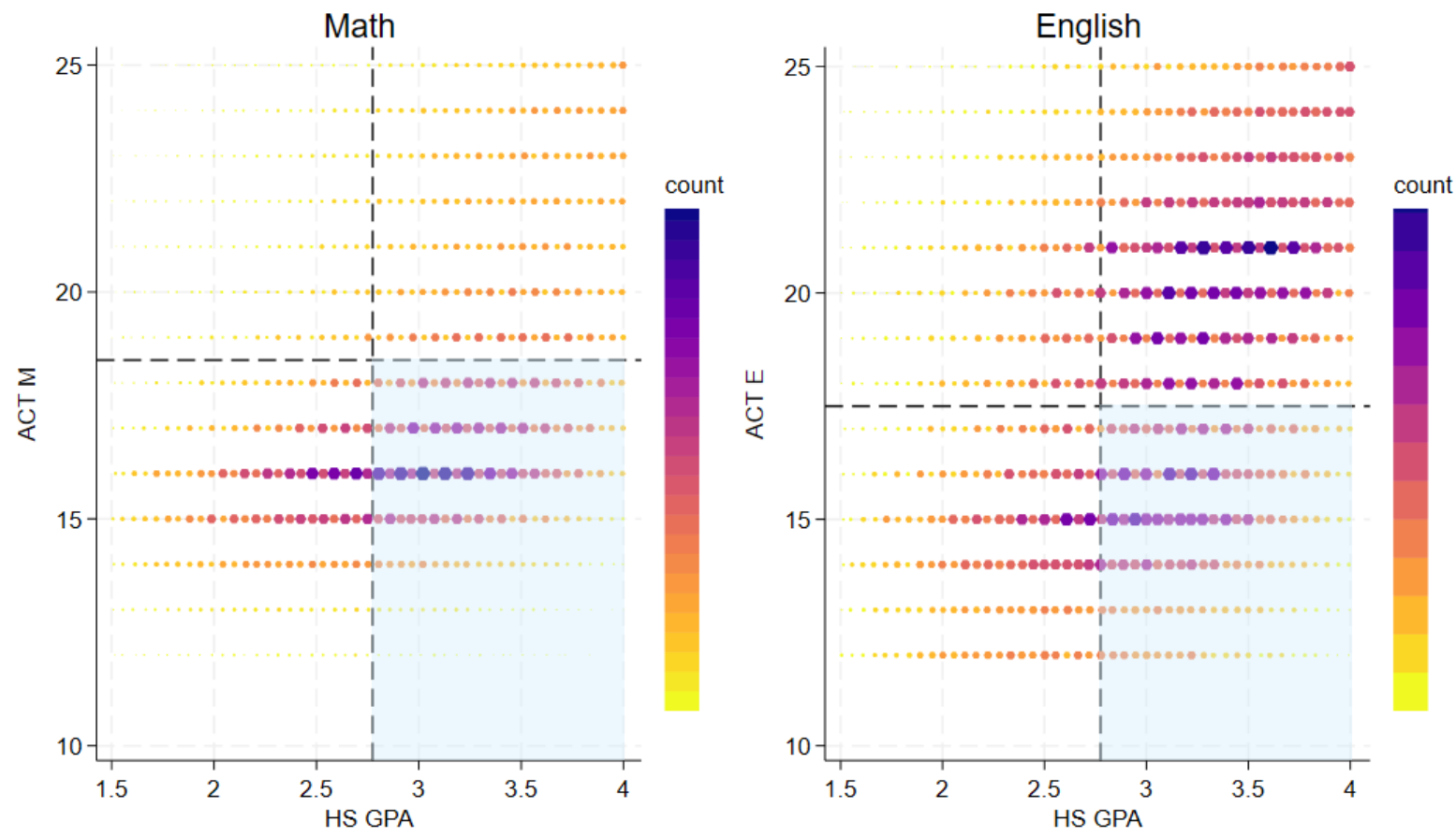
Challenges

- Traditional tests could misplace 25% or more of the students who could have succeeded in introductory college courses (Belfield & Crosta, 2012; Scott-Clayton et al., 2014).
- Racial minority students are overrepresented (75% vs. 64%) in DE (NCES, 2016)
- Under-placement adds unnecessary costs to students and institutions in both time and resources.

Adoption of MMA during the pandemic

- The shift was not new but accelerated (73% colleges using MMA as of 2023).
- Students can only be bumped up. No one gets bumped down.
- No fail and no zero HS grading policy proliferated.

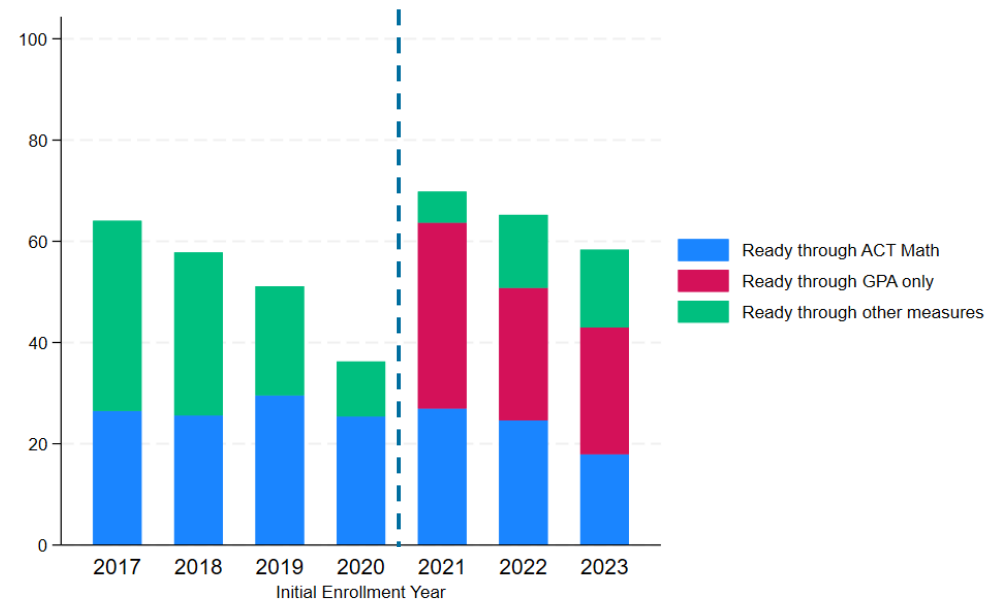
Potential scopes of change



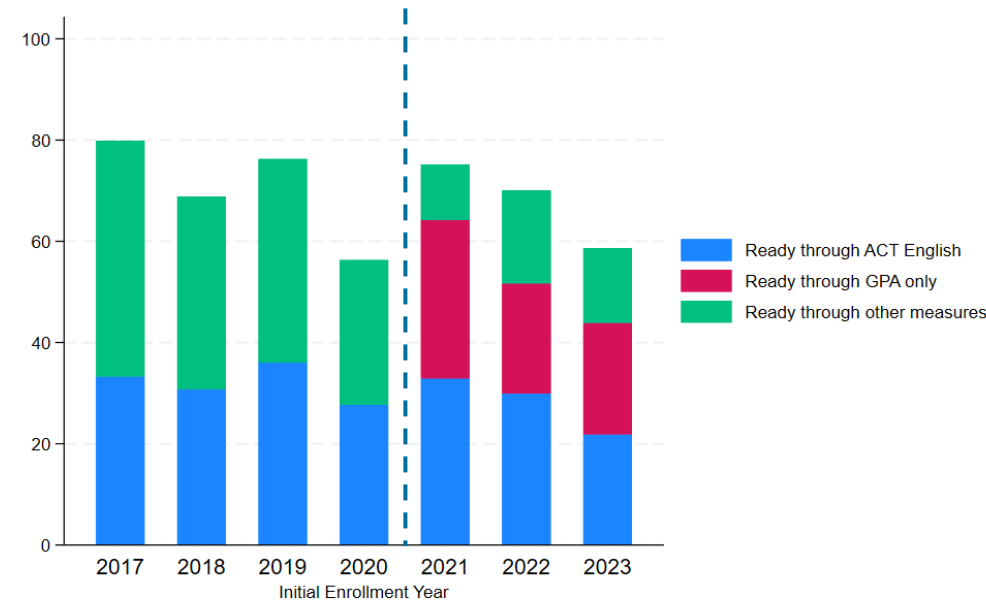
Note: Dashed vertical lines represent the HSGPA cutoffs (2.8), and dashed horizontal lines represent the ACT cutoffs (19 for math and 18 for English).

Percent of students deemed college ready, by measure

Math



English



Evidence on MMA

HS GPA could be helpful

- It reflects a wider range skills that are important for college success (Bailey & Jaggars, 2016)
- It is found to be more predictive of college completion than ACT scores (Allensworth & Clark, 2020)
- RCTs (SUNY, MN , WI) show positive effects for students in the bump up zone (Barnett et al., 2020, 2023; Cullinan & Biedzio, 2020). +15~19 pp in college math enrollment and +9~11 pp in completion.

Signs of challenge

- Negative effects for bumped down students. -16 pp in college math enrollment and -5 pp in completion.
- Lower odds of passing college courses among those who were bumped up and enrolled in those courses (67% vs. 74% in math and 81% in English).

Evidence on grading standards and grade inflation

Grading standards

- Students exposed to tougher standards tend to have higher achievement, college enrollment rates, and future earnings (e.g., Betts & Grogger, 2003; Hvidman & Sievertsen, 2021; Nordin et al., 2019).
- Lower grading standards tend to result in lower achievement (Bowden et al., 2023).

Behavioral responses to grade inflation

- Gains in GPA under lower standards are driven mostly by high-achieving students (Bowden et al., 2023).
- Increase in absences are driven entirely by low-achieving students.
- *Grade inflation does not necessarily lower the differentiating power of grades.*

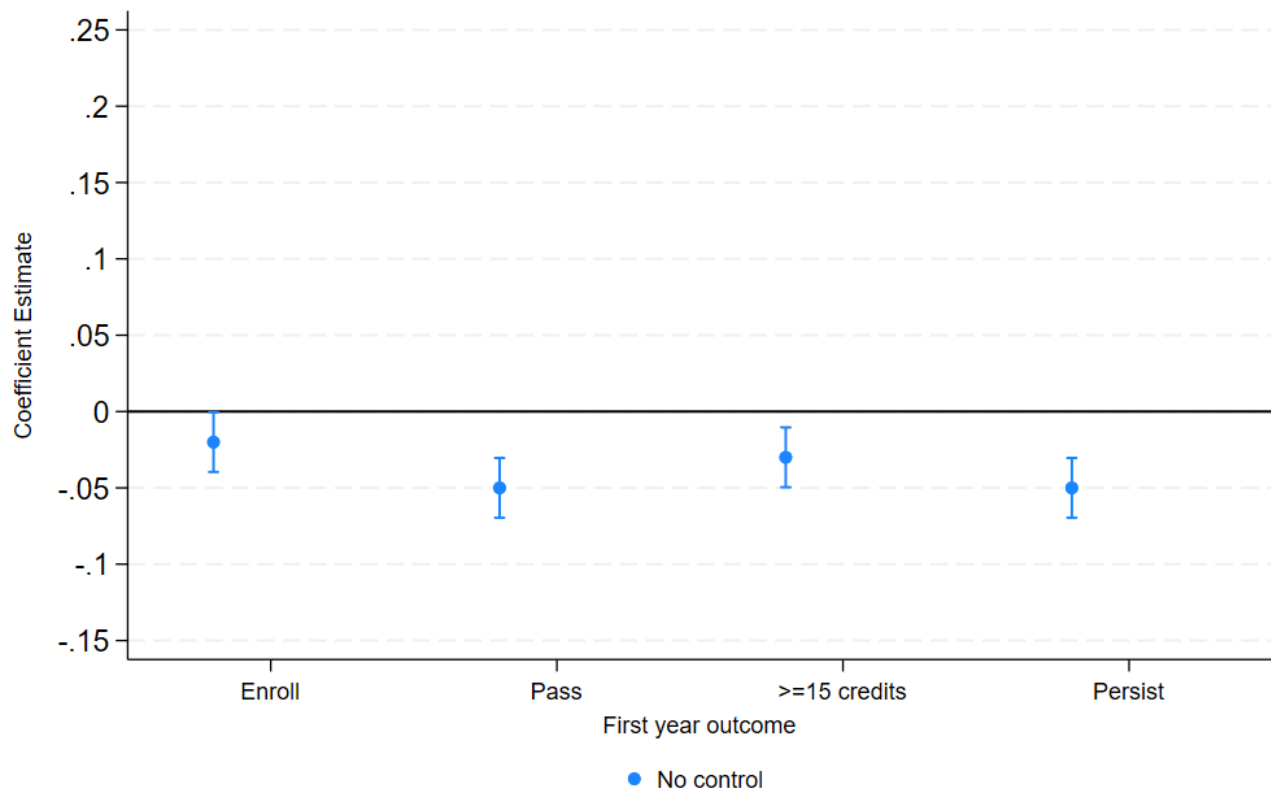
Research questions

1. Focusing on students in the bump up zone (treatment group = **Students $\geq$ HS GPA but $<$ ACT**)
 - A. Were students in the treatment group able to keep up with students $\geq$ ACT in 2020-21 and beyond?
 - B. Relative to the pre-waiver years, were students in the treatment group better off in the post-waiver period compared to students $\geq$ ACT?
 - C. Based on pre-waiver data, could students in the treatment group have benefitted from DE?
2. Predictive strength of HS GPA vs. the ACT and can the current MMA be improved

Data, sample, and measures

- **Setting:** All 16 colleges under KCTCS
- **Sample:** All first-time freshman enrolled in a degree program in 2016-17 through 2022-23, with cohorts starting in 2020-21 and later subject to placement policy waiver.
- **Data:** Postsecondary data merged with K-12 records. **ACT is mandatory** for all HS juniors.
- **Outcomes:** (a) enroll in college math or English in 1st year,
(b) complete college math or English in 1st year,
(c) taking at least 15 credits during the 1st year, and
(d) advance to the 2nd year.

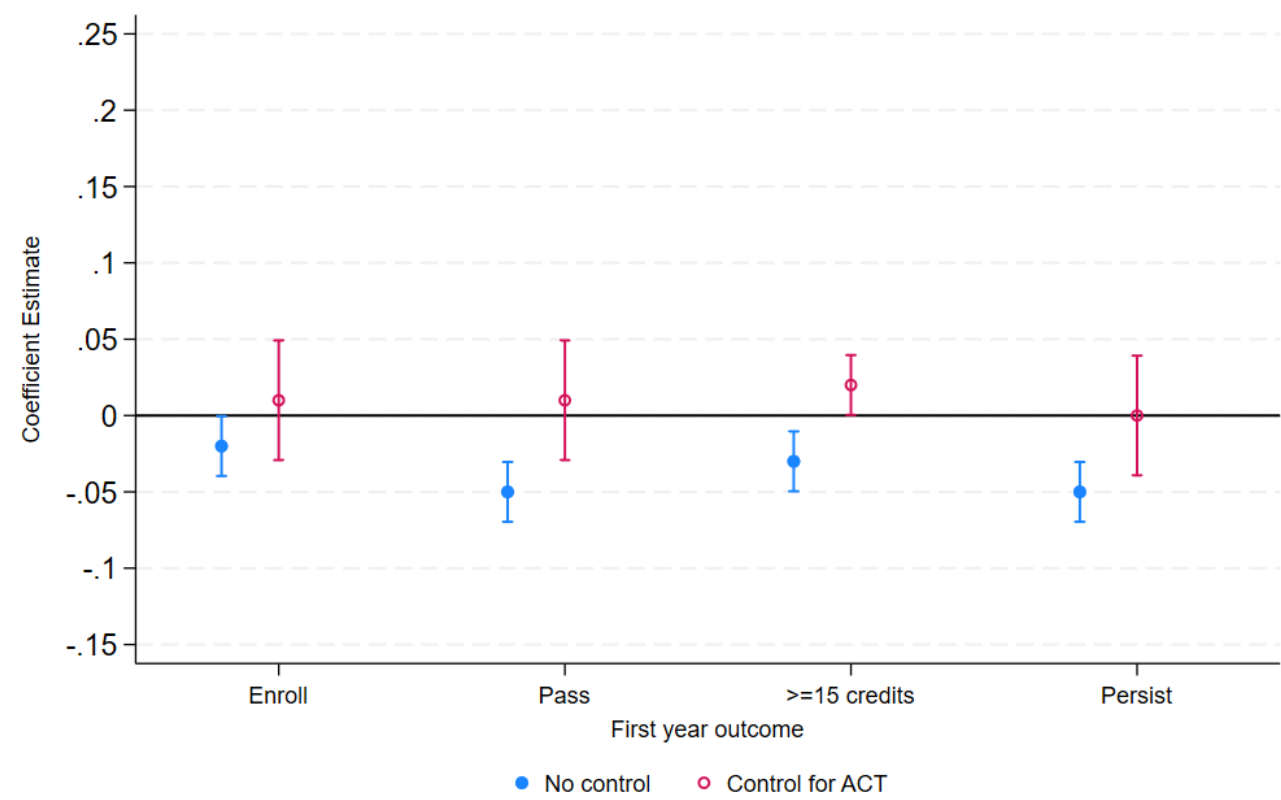
RQ1A. Could bumped up students keep up with students who met ACT benchmarks?



Math

- Students in the bump up zone had worse outcomes than students above the ACT math cutoff.

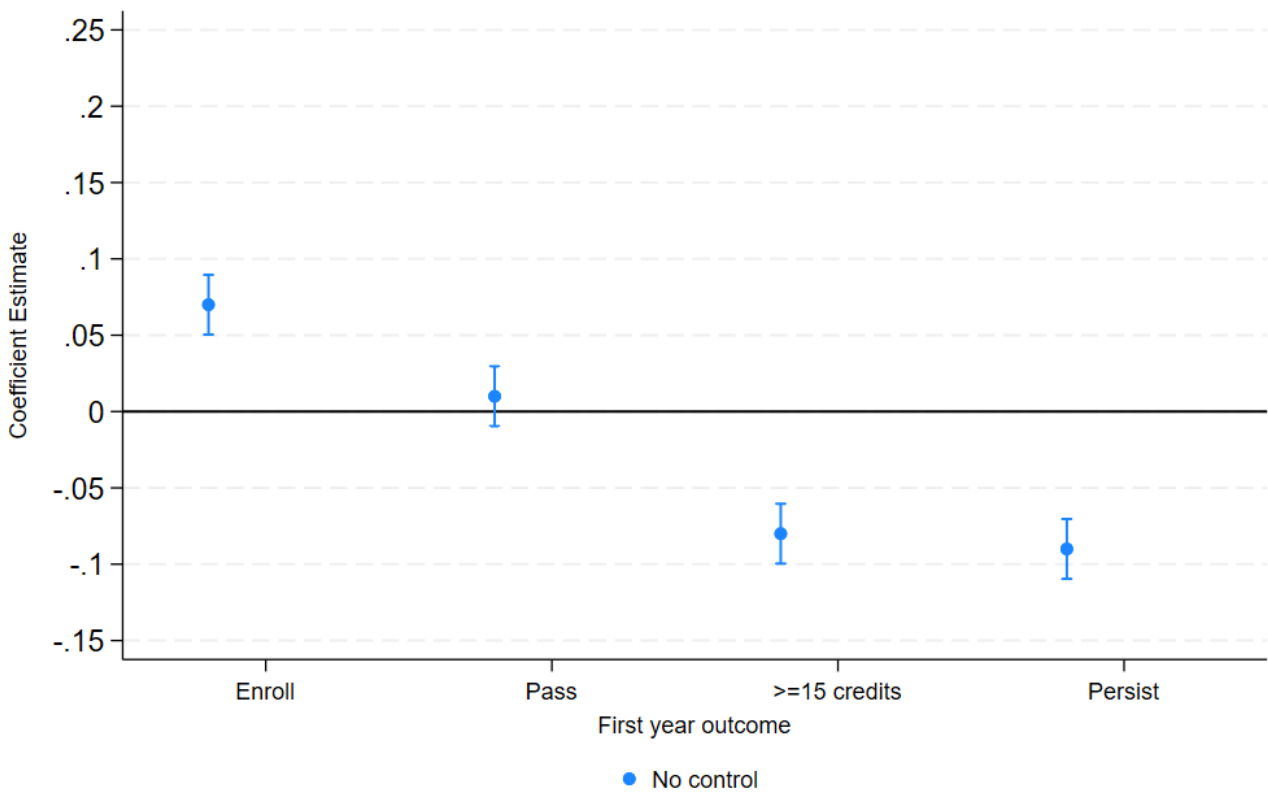
RQ1A. Could bumped up students keep up with students who met ACT benchmarks?



Math

- Students in the bump up zone had worse outcomes than students above the ACT math cutoff.
- Gaps in outcomes can be explained by lower pre-college skills as measured by the ACT.

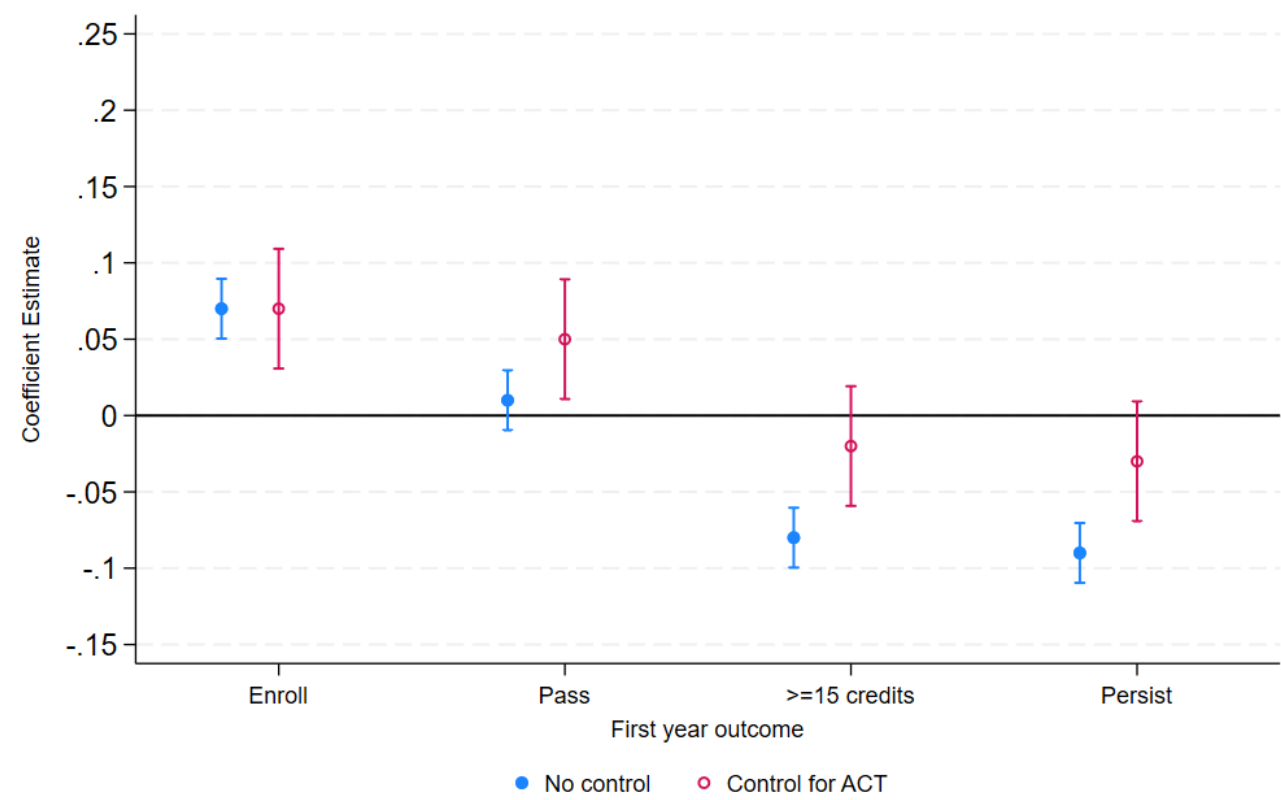
RQ1A. Could bumped up students keep up with students who met ACT benchmarks?



English

- Students in the bump up zone were more likely to take college English, but less likely to attempt ≥ 15 credits or advance to 2nd year.

RQ1A. Could bumped up students keep up with students who met ACT benchmarks?

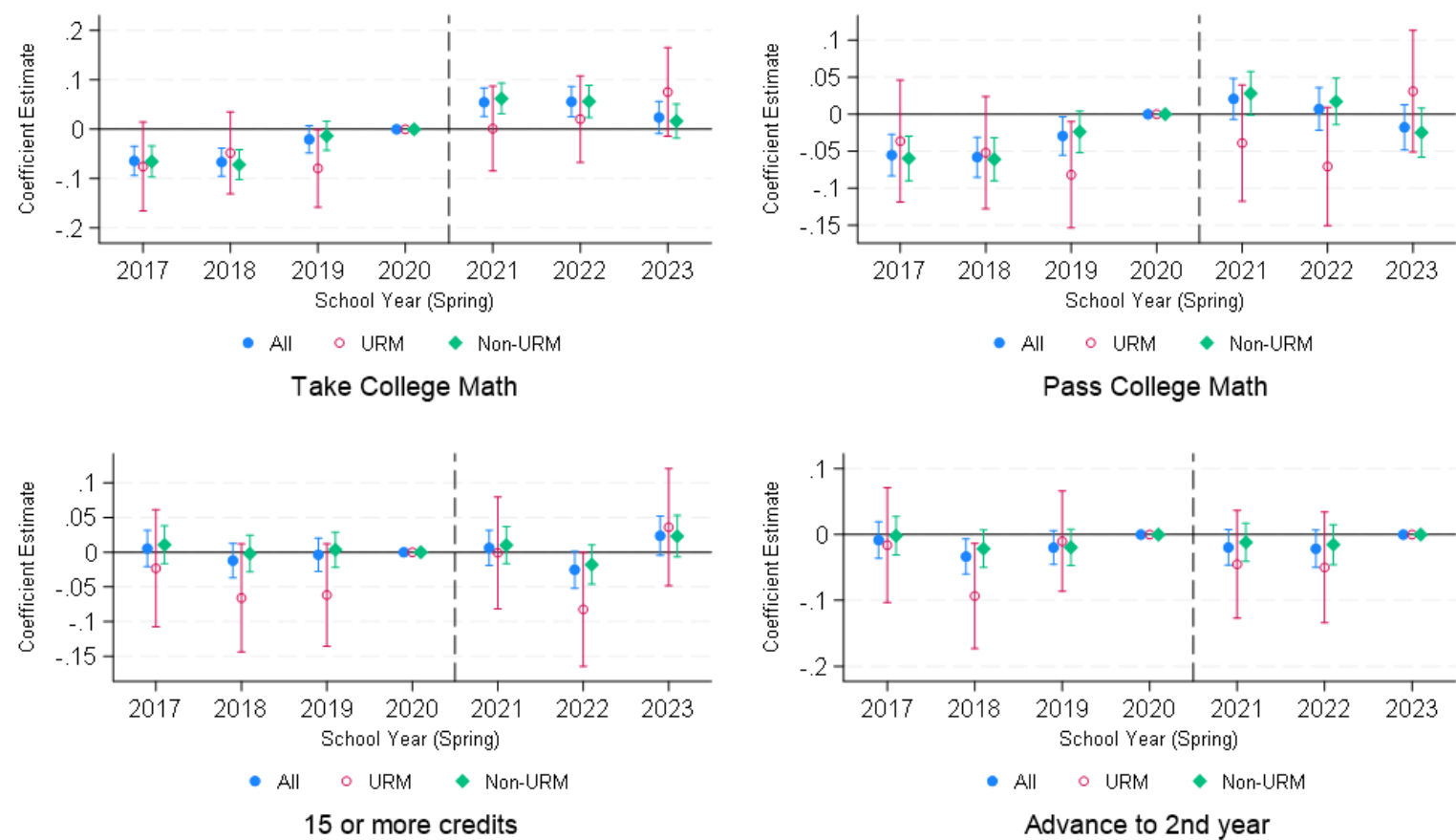


English

- Students in the bump up zone were more likely to take college English, but less likely to attempt ≥ 15 credits or advance to 2nd year.
- ACT scores can explain the gaps in credits attempted and persistence but not gateway English course enrollment.

RQ1B. Event study estimates suggest that changes had started before the waiver.

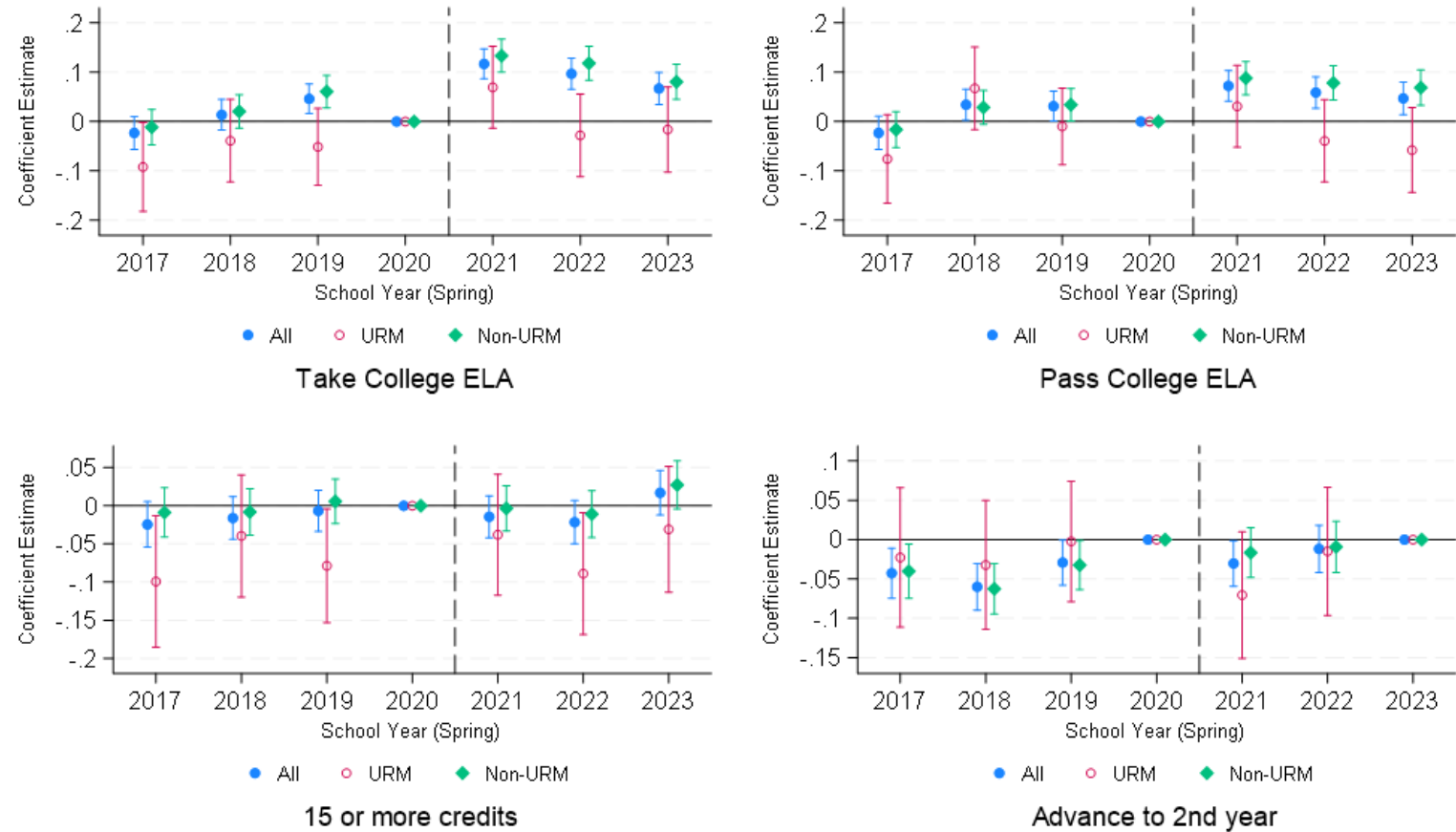
Math



Note: Coefficient estimates from a regression of a given outcome on a set of interactions between year fixed effects and an indicator for whether a student was in the bump up zone. Coefficients represent the impact of being in bump up zone relative to the final year before the policy change, 2020.

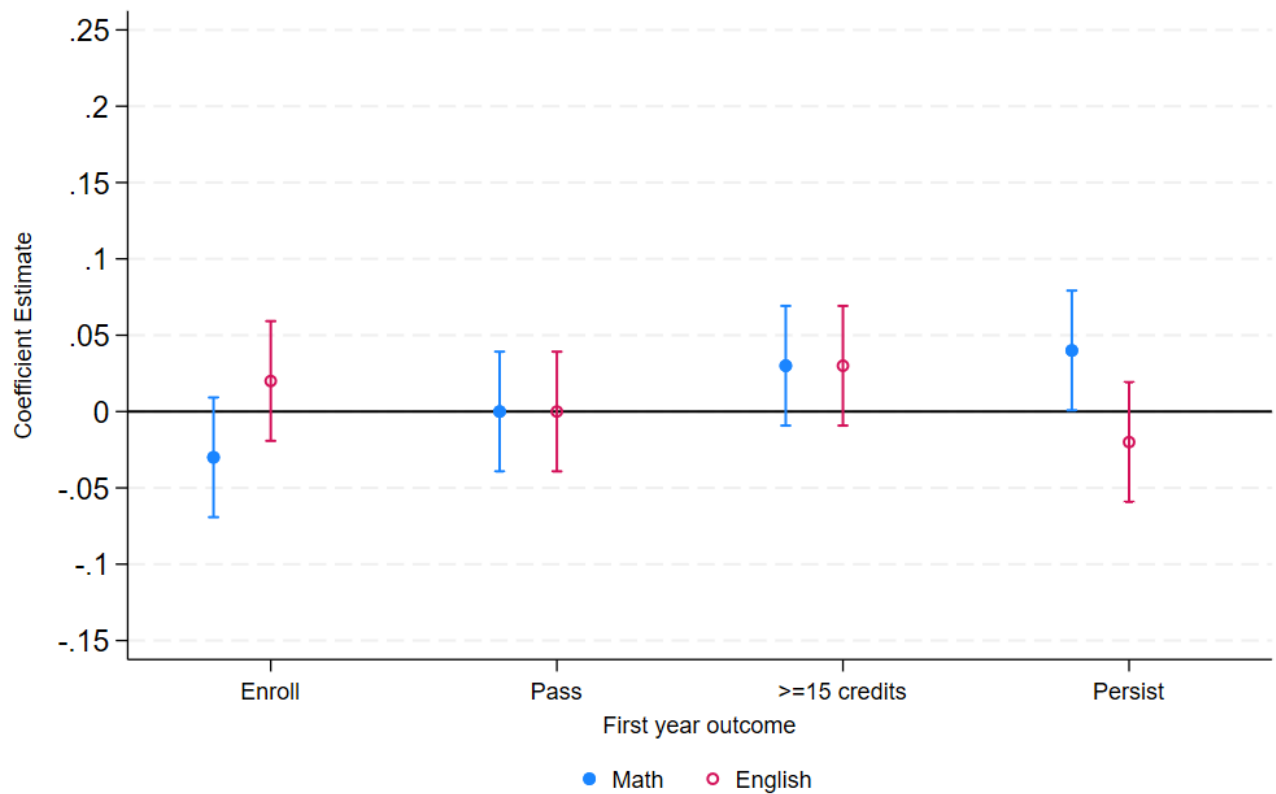
RQ1B. Event study estimates suggest that changes had started before the waiver.

English



Note: Coefficient estimates from a regression of a given outcome on a set of interactions between year fixed effects and an indicator for whether a student was in the bump up zone. Coefficients represent the impact of being in bump up zone relative to the final year before the policy change, 2020.

RQ1C: Could students in the bump up zone have benefitted from DE?

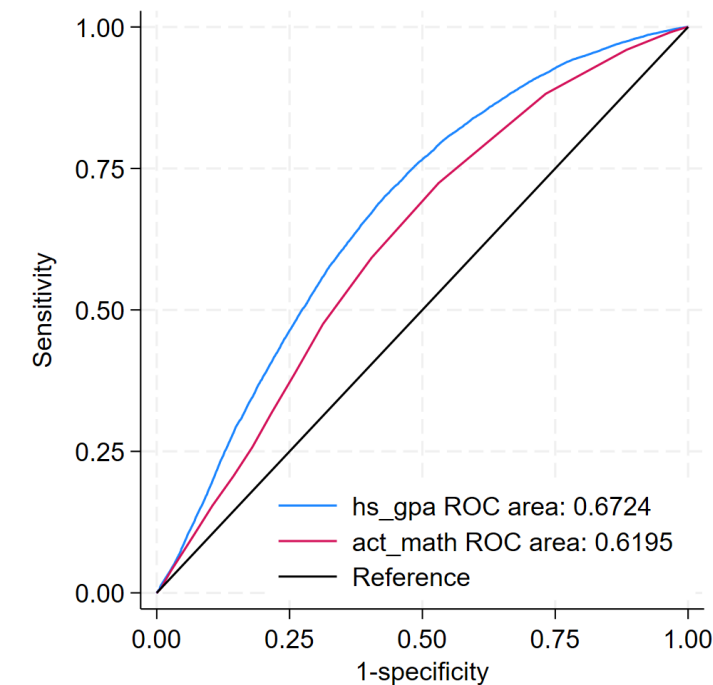


- No evidence suggests that students in the bump up zone benefitted from DE *as it was offered*. Except for persistence among students who received math DE.
- It does not necessarily rule out that those students *could have benefited* from DE had it been provided differently.

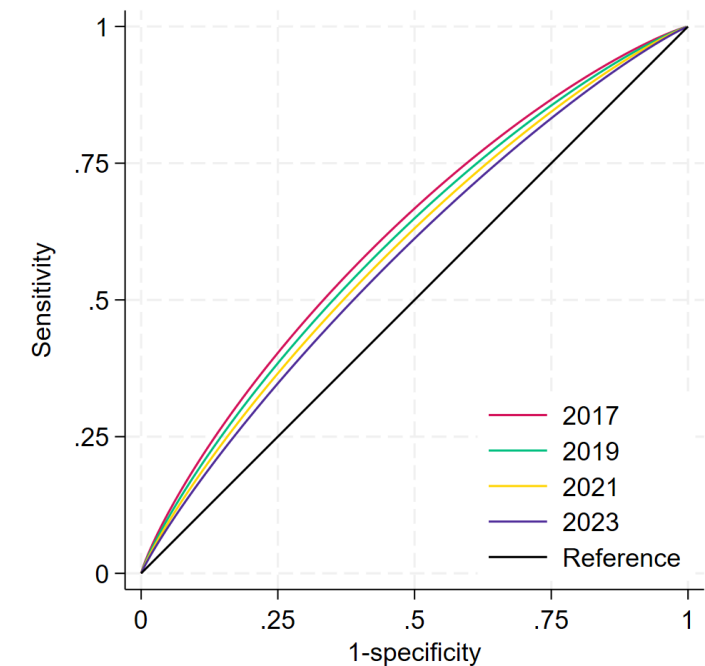
Note. Coefficients are estimated using regression discontinuity among pre-waiver students whose HS GPA would have exempted them from DE. ACT math is the running variable with a cutoff of 19. Bandwidth is set at 3. The model is estimated parametrically due to the discrete nature of ACT scores. Heteroskedasticity robust standard errors are used to calculate confidence intervals.

RQ2. HS GPA is a stronger predictor than ACT for passing gatekeeper math.

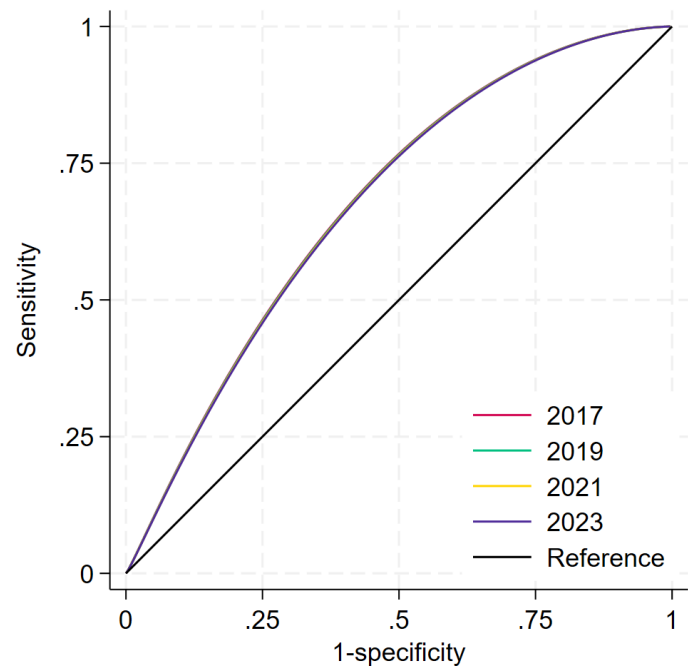
HS GPA is more predictive.



ACT math has become less predictive over time.



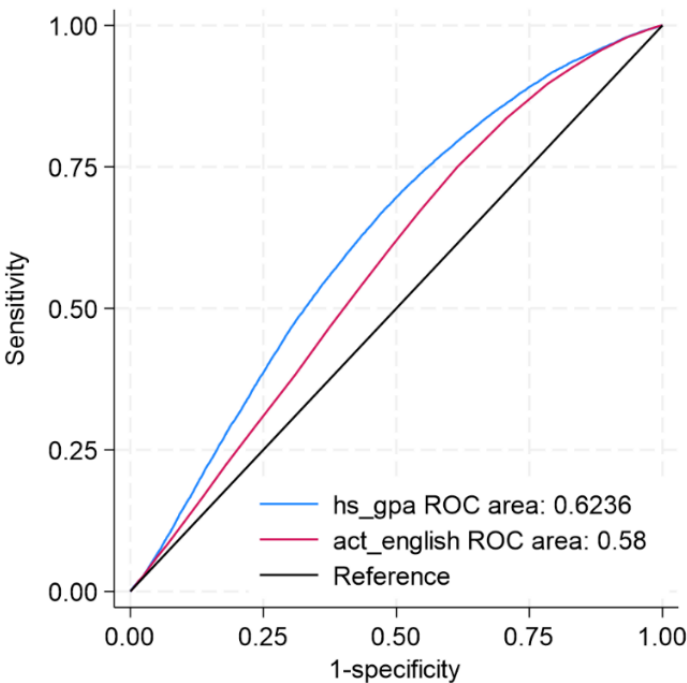
Predictive strength of HS GPA did not change over time.



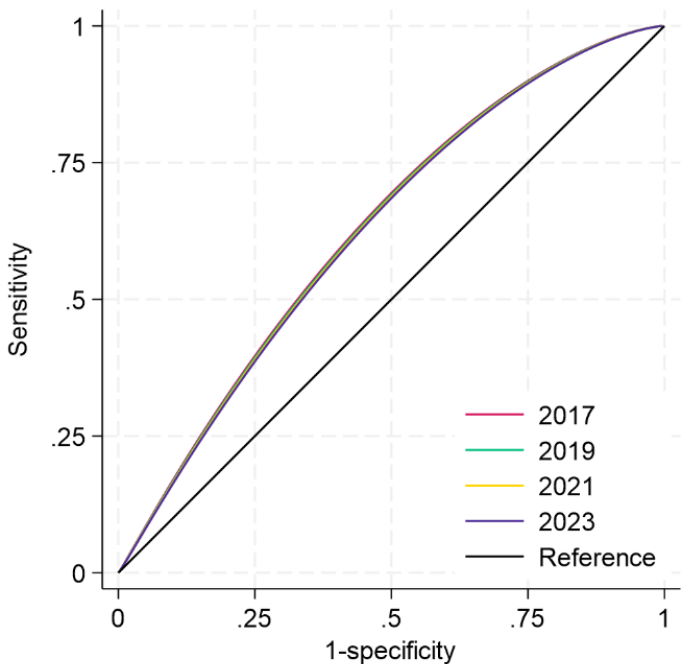
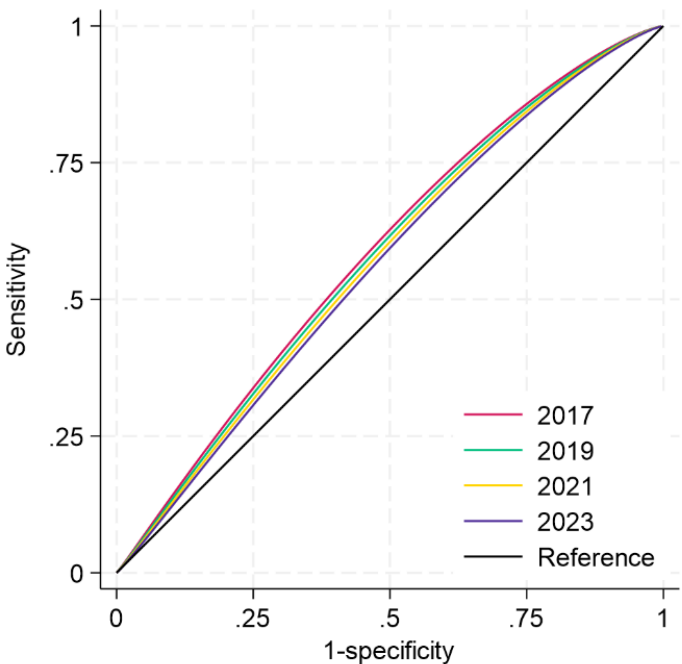
Note. Passing college math is the outcome in these examples. ROC (receiver operating characteristic) curve plots power (y-axis) against type-I errors (x-axis) using all possible cutoffs along the distribution of the predictor. The 45-degree reference line suggests prediction that is as good as random guess. Larger area under the curve (AUC) suggests stronger predictive power.

RQ2. HS GPA is a stronger predictor than ACT for passing gatekeeper English.

HS GPA is more predictive.



ACT English has become less predictive over time. Predictive strength of HS GPA did not change over time.



Note. Passing college English is the outcome in these examples. ROC (receiver operating characteristic) curve plots power (y-axis) against type-I errors (x-axis) using all possible cutoffs along the distribution of the predictor. The 45-degree reference line suggests prediction that is as good as random guess. Larger area under the curve (AUC) suggests stronger predictive power.

RQ2. Prediction accuracy of aggregate MMAs

- Overall accuracy is 69% for math and 73% for English using a linear combination of overall HS GPA and ACT subject score.
- Accuracy changes little when using subject-specific GPA and adding other ACT subject scores.
- Accuracy changes little when using logistic regression or XGBoost.

Summary

- Students designated as college ready based on HS GPA alone had **weaker** first-year outcomes than students who met the ACT benchmarks, except for college English enrollment.
- The **gaps were narrowing** in college course enrollment and completion, but these gains likely reflect the combined effect of multiple aspects of DE reforms.
- Students designated as college ready in math based on HS GPA alone could have improved persistence with extra support, but in general students in the bump up zone **did not benefit from additional DE**.
- HS GPA is a stronger predictor for gateway course completion than ACT.
- Subject-specific HS GPA, additional subject ACT scores, and estimation methods result in minimal changes in predictive accuracy.

Disrupted Trajectories: College Enrollment and Employment among High School Graduates through the COVID-19 Pandemic

CALDER Conference: March 12, 2025

Elc Estrera

Wake County Public School System

Steven W. Hemelt

UNC-Chapel Hill / CALDER

Matthew Lenard

Florida State University

Slides for this paper will be added when
results are public and ready to share

Session III: The Future of Annual Testing—Roundtable

Panelists

- **Denise Forte**, President and CEO, The Education Trust
- **Aimee Rogstad Guidera**, Secretary of Education for the Commonwealth of Virginia
- **Michael J. Petrilli**, President, Thomas B. Fordham Institute

Moderator

- **Dan Goldhaber**, CALDER/AIR

Agreement on the Need for Change But Not the Path Forward

The status quo testing system produces results that are too broad and late to be useful to educators and families. What should we do instead?



FORUM STANDARDS, TESTING, AND ACCOUNTABILITY

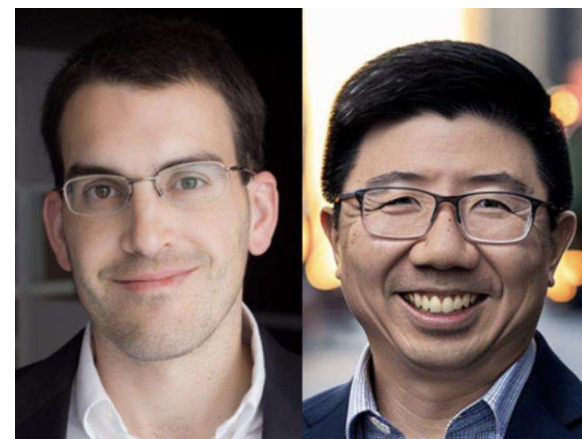
Which of the Following Approaches to State Testing Works for U.S. Schools?

“States could reduce the amount of testing by borrowing the **sampling approach** used by pollsters.”



Lynn Olson & Tom Toch

“Our preferred vision for state tests is to make them more akin to rapid Covid tests or new pregnancy tests that promise **ultra-fast results.**”



Chad Aldeman & Dale Chu

Testing Purposes & Testing Alternatives

Four Purposes of Educational Tests

Groups (e.g., classrooms, schools, states, countries)	The Monitoring Quadrant (e.g., NAEP, TIMSS, PISA)	The Accountability Quadrant (e.g., teacher value added, state tests)
	The Classroom Quadrant (e.g., formative, feedback)	The Selection Quadrant (e.g., admission, awards, remediation, certification)
Individuals	Low Stakes	High Stakes

Source: Ho (2023). Towards assessment literacy: Questions to ask about educational tests. Harvard Graduate School of Education

ESEA testing status quo

Annual testing of all students in grades 3-8 and once in high school in math and English language arts. Science testing once in elementary, middle, and high school.

Some testing alternatives

- **Sample students:** E.g., test a sample students, or test everyone but only once in elementary, middle, and high school.
- **Sample content:** E.g., spread testing items across students (individuals take part of the total test)—*matrix sampling*.
- **Faster, cheaper tests:** E.g., quicker release of results to the public; streamline verification procedures.
- **Through-year assessments:** E.g., give multiple, shorter tests during the school year to inform instruction and produce an end-of-year score.
- **Performance-based and/or competency-based assessments:** E.g., have students demonstrate what they know via a task or creating a product; allow students to demonstrate competences as they move at their own pace.

That's a Wrap. Thank You!

Visit us at caldercenter.org

Follow us on Bluesky at [@caldercenter.bsky.social](https://bsky.app/profile/caldercenter.org) and on X at [@caldercenter](https://twitter.com/caldercenter)

CALDERCENTER.ORG

